RIJKSUNIVERSITEIT GRONINGEN

The MemGen structural classification of secondary transporters
A membrane topology screening approach

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag13 juli 2012
om 12.45 uur

door

Ramon ter Horst

geboren op 31 maart 1973
te Enschede

Promotor:	Prof. dr. A.J.M. Driessen
Copromotor:	Dr. J.S. Lolkema
Beoordelingscommissie:	Prof. dr. E.J. Boekema
	Prof. dr. J. Kok
	Prof. dr. D.J. Slotboom

CONTENTS

Chapter 1

General Introduction - Secondary transport proteins: Classification and membrane topology prediction

All living cells are surrounded by at least one biological membrane that forms a hydrophobic barrier between the interior of the cell and its environment. The uptake of nutrients, excretion of metabolic end products, removal of toxic compounds, maintaining osmolarity and internal pH require designated transport proteins that are embedded in the membrane. These integral membrane proteins are essential for selective solute transport and allow the cell to maintain homeostasis to ensure optimal conditions for metabolic processes. The importance of membrane proteins is further illustrated by the fact that about 20-30% of all genes in most genomes encode membrane proteins [1,2].

Transport proteins can be classified on the basis of their mode of transport and energy coupling mechanism. Four major classes can be distinguished:

- Channels, that allow selective solute transport down the concentration gradient by means of facilitated diffusion;
- Primary transporters (ABC transporters), which couple hydrolysis of ATP to the uptake or excretion of substrates;
- Secondary transporters, which use the free energy stored in electrochemical gradients across the membrane;
- Group translocation systems, which couple translocation to chemical modification of the substrates.

The above-mentioned classification will be discussed in detail in section 1.

1. Functional-phylogenetic classification of transport proteins: the TC system

The Transport Classification (TC) system classifies protein families into transporter classes and subclasses [3] that distinguish functionally distinct types of transporters. Classes and subclasses are further subdivided in families and subfamilies that provide a phylogenetic basis for the classification. The TC system is therefore a functional-phylogenetic system. Over 500 protein families are currently found in the TC system, grouped according to transporter class and subclass, as presented in table 1 [4]. The classification of transport proteins is based on five criteria, and each of these criteria corresponds to one of the five digits (D) or letters (L) within the TC number. D1 and L2 correspond to the transporter class and subclass, respectively, and D3, D4 and D5 refer to (super)family, subfamily and (range of) substrate(s) transported. Any two transporters in the same subfamily of a transporter family that transports the same substrate(s) using the same mechanism are given the same TC number, regardless of whether they are orthologs (i.e., arose in different organisms by speciation) or paralogs (i.e., arose within a single organism by gene duplication) [3]).

Table 1.

Classes and subclasses of transport systems included in TCDB (from [19])

1.	Channels/pores
	1.A α-Type channels
	1.B β-Barrel porins
	1.C Pore-forming toxins (proteins and peptides)
	1.D Non-ribosomally synthesized channels
	1.E Holins
	1.F Vesicle fusion pores
	1.G Paracellular channels
2.	Electrochemical potential-driven transporters
	2.A Porters (uniporters, symporters, antiporters)
	2.B Nonribosomally synthesized porters
	2.C Ion-gradient-driven energizers
3.	Primary active transporters
	3.A P-P-bond-hydrolysis-driven transporters
	3.B Decarboxylation-driven transporters
	3.C Methyltransfer-driven transporters
	3.D Oxidoreduction-driven transporters
	3.E Light absorption-driven transporters
4.	Group translocators
	4.A Phosphotransfer-driven group translocators
	4.B Nicotinamide ribonucleoside uptake transporters
	4.C Acyl CoA ligase-coupled transporters
5.	Transport electron carriers
	5.A Transmembrane 2-electron transfer carriers
	5.B Transmembrane 1-electron transfer carriers
8.	Accessory factors involved in transport
	8.A Auxiliary transport proteins
	8.B Ribosomally synthesized protein/peptide toxins that target channels and carriers
	8.C Non-ribosomally synthesized toxins that target channels and carriers
9.	Incompletely characterized transport systems
	9.A Recognized transporters of unknown biochemical mechanism
	9.B Putative transport proteins
	9.C Functionally characterized transporters lacking identified sequences

The primary level of classification in the TC system is mode of transport and energy-coupling source, and 7 classes are currently recognized. The well-defined classes 1 to 5 correspond to

Channels or pores, Electrochemical potential-driven transporters, Primary active transporters, Group translocators and Transport electron carriers, respectively. Classes 6 and 7 are not yet defined. Class 8 is reserved for accessory/auxiliary transport proteins and class 9 corresponds to (families of) transporters that are not completely characterized.

The Channel-Type Facilitators category, class 1 of the TC system, includes almost 300 families grouped in 4 different subclasses, the α-Type channels (1.A), the β-Barrel porins (1.B), the Pore-forming toxins (1.C) and the Non-ribosomally synthesized channels (1.D). Proteins of this class are found in all kingdoms of life, from bacteria to higher eukaryotes. Members of this class are transmembrane channels or pores that catalyze facilitated diffusion by an energy independent process. They do not exhibit stereospecificity, but may be specific for a particular molecular species or class of molecules. The largest group in this class is the Voltage-gated Ion Channel (VIC) superfamily.

The Electrochemical Potential-Driven Transporters category corresponds to class 2 of the TC system. This class is divided in 3 subclasses that include 116 secondary transporter (super)families, i.e. Porters (2.A; uniporters, symporters, antiporters), Non-ribosomally synthesized porters (2.B) and Ion gradient-driven energizers (2.C). Porters or Secondary transporters constitute one of the largest functional categories. Transporters found in class 2 are universal to all biological cells and play a role in many biological processes. Sequence analysis revealed the existence of a few large superfamilies in this class, e.g. the major facilitator superfamily (MFS) [5], the amino acid/polyamine/organocation superamily (APC) [6], the ion transporter superfamily (IT) [7], and many smaller (super) families. Secondary transporters catalyze the translocation of substrates across membranes driven by (electro) chemical gradients of substrates and co-ions. They are the main subject of this thesis.

The Primary Active Transporters are grouped in category 3. TC system class 3 is divided in 5 subclasses, the Diphosphate bond hydrolysis- (3.A), the Decarboxylation- (3.B), the Methyl transfer- (3.C), the Oxidoreduction- (3.D) and the Light absorption-driven transporters (3.E). The Diphosphate bond hydrolysis- and the Oxidoreduction-driven transporters, represented by 30 families (e.g. (3.A) ATP-binding Cassette (ABC) Superfamily, and (3.D) H^+ or Na^+-translocating NADH Dehydrogenase(NDH) family), are found universally in all living organisms. Only a few families are restricted to one or another domain of life. The decarboxylation-driven transporters (3.B Na+-Transporting Carboxilic Acid Decarboxylase (NaT DC) family) and the methyl

transfer-driven transporters (3.C Na+-transporting Methyltatrahydromethanopterin:Coenzyme M Methyltransferase (NaT-MMM) family) are thought to be restricted to prokaryotes and archaea, respectively. The subclass Light absorption-driven transporters (3.E) contains two families, the Ion-translocating Microbial Rhodopsin (MR) family and the Photosynthetic Reaction Center (PRC) family. The MR family is found in archaea and eukaryotes, and the PRC family only in bacteria and chloroplasts of eukaryotes. Transporters in class 3 of the TC system use a primary source of energy (chemical, or solar) to drive transport of a solute against a concentration gradient.

Class 4 comprises the Group Translocators. The Group Translocators are divided in 3 putative subclasses, e.g. the Phosphotransfer-driven group translocators (4.A), the Nicotineamide ribonucleoside uptake transporters (4.B) and the Acyl CoA ligase-coupled transporters (4.C). A role in group translocation of the latter two types of transporters is not fully accepted [8]. The bacterial sugar transporters of the phosphoenolpyruvate-dependent phosphotransferase system (PTS) are the best-characterized group translocators included in class 4. No transporters of the PTS subclass have been identified either in archaea or in eukaryotes. Group translocation involves chemical modification of the substrate during the transport process. The coupling of transport and substrate modification may be very tight and transport is inhibited without substrate modification. Alternatively, in a more loosely coupled substrate modification and translocation process, transport without modification may still occur, although the normal process involves concurrent substrate modification.

The Transmembrane Electron carriers, the newly emerged class 5, includes 9 transporter families grouped in 2 different subclasses, the Transmembrane 2 (5.A) and the Transmembrane 1 electron transport carriers (5.B). Transporters found in TC class 5 have been identified in bacteria and archaea. Members of subclasses 5.A and 5.B, transfer electron pairs and single electrons across the membrane from an electron (pair) donor on one side of the membrane to an electron(pair) acceptor on the other side, respectively. A few characterized proteins in this class translocate cytoplasmic electrons to oxidized disulfide-containing proteins in the periplasm [9,10,11].

Assignments for classes 6 and 7 have not yet been made. These classes are reserved for novel types of transporters that do not fall into classes 1 to 5.

Classes 1-5 include all of the currently recognized, well characterized transport systems with established energy coupling mechanisms. Many transporters are insufficiently characterized to be

included in one of these classes. They are given a classification number (8 or 9) that indicates their degree of incomplete characterization. Class 8 has been reserved for accessory/auxiliary transport proteins. Included proteins facilitate, but do not directly participate in transmembrane transport. They may play a role in f.i. the coupling of energy and transport, or complex formation, or serve a biogenic, stability or regulatory function. Class 9 corresponds to transporter families of which at least one member of each family has clearly been shown to function as a transporter, but with unknown mode of transport or energy-coupling mechanism.
Families and subfamilies provide a phylogenetic basis to the classification, reflecting evolutionary relationships of transport proteins. In the TC system, remote relationships are evidenced by the existence of several superfamilies that group distantly related families, like the Major Facilitator Superfamily (MFS), the acid/polyamine/organocation (APC) superfamily and the ion transporter (IT) superfamily. Crystal structures of two proteins from different families in the MFS, the lactose transporter LacY [12] and the glycerol-P/Pi exchanger GlpT [13] of *E. coli*, revealed very similar structures of these distantly related proteins. In another approach, involving hydropathy profile alignment, more distant evolutionary relationships are established [14]. This approach is based on the fact that during evolution structure is much better conserved then amino acid sequence. It follows that proteins grouped in a family based on sequence identity share a similar fold as well. Moreover, comparing the structures of proteins allows for detection of more distant evolutionary relationships than sequence alignment alone.

2. Structural classification of secondary transport proteins by hydropathy profile alignment

Secondary transporters are integral membrane proteins commonly encoded by a single gene. They consist of a bundle of α-helices that are more or less perpendicular to the plane of the membrane. Despite this simple architecture, their phylogenetic diversity is enormous as evidenced by the more than 100 gene (super)families in the Electrochemical potential-driven transporters section of the Transport Classification system (TC system) [3]. Most likely, the genetic diversity is a consequence of divergent evolution and many different families may represent a similar fold and translocation mechanism. In recent years high resolution X-ray structures have been presented of secondary transporters from different gene families. Some of them showed the same core structure while no significant sequence similarity could be identified

between the transporters. The families include the Neurotransmitter:Sodium Symporter (2.A.22 NSS) family (structure of LeuT) [15], the Solute:Sodium Symporter (2.A.21 SSS) family (vSGLT) [16], the Nucleobase:Cation Symporter-1 (2.A.39 NCS1) family (Mhp1) [17], the Betaine/Carnitine/Choline Transporter (2.A.15 BCCT) family (BetP) [18] and the Amino Acid-Polyamine-Organocation (2.A.3 APC) family (AdiC, ApcT) [19,20]. The structural similarity between transporters of the NSS, SSS, NCS1 and APC families was predicted before by the MemGen classification system [21], in which hydropathy profile analysis is used as a tool to group families of membrane proteins in structural classes.

Membrane proteins consist of bundles of membrane spanning α-helices connected by hydrophilic loops. An α-helix of about 20 amino acids with high hydrophobicity is able to span the hydrophobic core of the phospholipid bilayer, while the interconnecting loops that contact the water phase have low average hydrophobicities. The alternating regions of high and low hydrophobicity give rise to a hydropathy profile characteristic to the membrane protein, and provides a fingerprint of the structure of the protein [22]. An example of a membrane protein hydropathy profile is shown in Fig. 1.

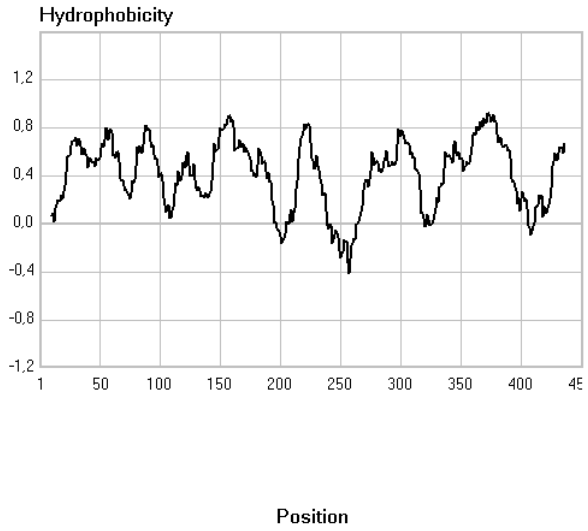


Fig. 1. Hydropathy profile of a membrane protein

Although a high level of structural detail cannot be deduced from the hydropathy profile directly, they can be used to discriminate between different structures, or to show similarities between apparent non-homologous proteins or protein families by comparison of the hydropathy profiles.

MemGen uses hydropathy profiles as fingerprints of the global fold of membrane proteins in order to detect structural similarity. First, an average family hydropathy profile is calculated from the multiple sequence alignment of members of a family, with unambiguous similarity at the amino acid level. The family hydropathy profile is considered a more accurate fingerprint of the structure of the members of the family because it reduces the noise in the profile introduced by the individual amino acid sequences. Then, the optimal alignment of two family profiles is computed without making use of the underlying sequences [23] to detect similarity of the (family) hydropathy profiles. The alignment procedure is very similar to the approach used in amino acid sequence alignment that tries to mimic an evolutionary pathway by which one sequence is converted into the other by replacing amino acid residues and introducing gaps and deletions while minimizing the cost of conversion [23,24]. Similarly, one hydropathy profile is converted to the other while scoring the cost as the difference in hydrophobicity at each position. The technique can be used for similarity detection of protein families that share only marginal or no sequence identity whatsoever. A quantitative assessment of (dis)similarity of the optimally aligned family hydropathy profiles was found by comparing the difference between these profiles (PDS; Profile Difference Score), and the averaged difference between the individual profiles and the family profile in each family (SDS; Structure Divergence Score). The PDS is defined as the average square of the differences between two hydropathy profiles at each position in the alignment. The SDS quantifies the divergence within a family by averaging the square of the differences between the family profile and the individual profiles at each position. The comparison between PDS and SDS is expressed in a similarity test, the S-test, which is defined as the ratio of the square of the PDS and the average square of the SDSs of the two families. S-test values of 1 or lower signify similar structures [23,25].

The method described above led to the MemGen classification, in which families of membrane proteins are grouped in structural classes according to family hydropathy profile identities. The classification identified evolutionary relationships between secondary transporter families of which the members were apparently unrelated on the basis of sequence identity. In a nutshell, family profiles were used to screen for structurally similar protein families with S-test values lower than a certain threshold value. The selected alignments were visually inspected for false positives, which were removed from the list, and the similar structures were pooled in 4 structural classes. Further optimization of the classification involved re-evaluation, and re-assignment or

rejection of entries (and their paralogues) that were present in more than one structural class. When the method was introduced in 1998, 8 families of secondary transporters were classified into 4 different structural classes, termed ST[1], ST[2], ST[3] and [ST4] [23,24]. Later, when available sequence databases were screened, many more families were assigned to these structural classes [21,26,27]. Within a structural class, two levels of evolutionary distance were recognized [26]. Subfamilies contain sequences that show significant overall sequence identity in a multiple sequence alignment, and families that contain one or more subfamilies. Members of different subfamilies in one family show significant sequence identity in a local sequence alignment as produced by the BLAST algorithm [28]. For instance, structural class ST[3] contains over 30 families and approximately 90 subfamilies. More than 50 gene (super)families in the Electrochemical potential-driven transporters section of the Transport Classification system (TC system) [3] are classified in one of the 4 currently recognized classes of the MemGen structural classification system. Structural class ST[1] in the MemGen classification system largely matches with the Major Facilitator Superfamily (MFS) in the TC system [23,24]. Class ST[2] contains the Amino Acid-Polyamine-Organocation (APC) superfamily, the Solute:Sodium Symporter (SSS) and Neurotransmitter:Sodium Symporter (NSS) families, and a number of smaller transporter families. Structural class ST[3] is a well studied class of the MemGen classification system [26,27,29,30,31]. Thirty-six families were classified in ST[3], among which the Ion Transporter (IT) superfamily members, the 2-Hydroxycarboxylate Transporter (2-HCT) Family, and the Glutamate:Na^+ Symporter (ESS) Family [23,26]. Finally, structural class ST[4] contains a single family of transporters, the Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) Family, including glutamate, neutral amino acids and C_4-dixarboxylate transporters. Fig. 2 shows a screen copy of the MemGen structural classification of secondary transport proteins, with structural class ST[3] collapsed to show the protein (sub)families included.

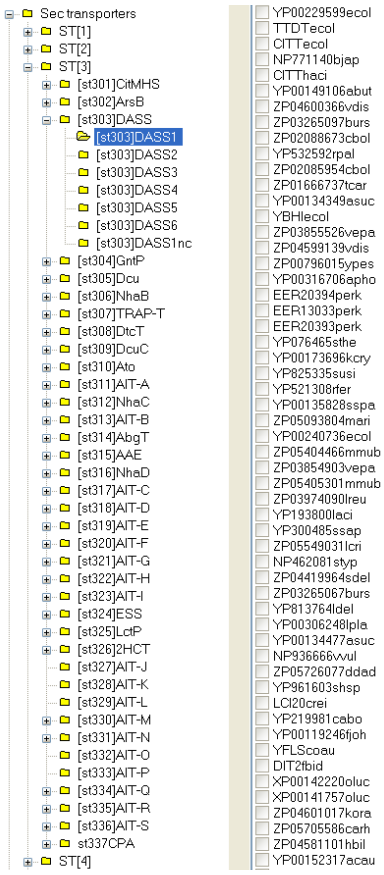


Fig. 2. MemGen structural classification of secondary transporters. Secondary transport protein families (and subfamilies) are classified into one of four different structural classes.

The recently reported 3D crystal structures of secondary transporters supported the MemGen classification. They showed that transporters from different classes (LacY and GlpT in ST[1] [12,13], LeuT in ST[2] [15], and Glt_{Ph} in ST[4] [32]) clearly represented different structures. Support for the MemGen classification within a structural class was first provided by high resolution structures of LeuT, a member of the NSS family [15] and vSGLT of the SSS family [16], and the later reported structures of Mhp1 [17], BetP [18] and, AdiC and ApcT [19,20], members of the NCS1, BCCT and APC families, respectively. While unrelated in sequence (except for AdiC and ApcT), these transporters are grouped in ST[2] and share a core structure with the same global fold. At a lower level of resolution, the structural similarity has been demonstrated for families in class ST[3] [21]. A major feature of the MemGen classification system is that all proteins in the different families in one structural class share the same global fold. Although the MemGen classification system is not a membrane topology prediction method in itself, knowing the topology of one protein in a structural class infers knowing them all [33].

3. Bioinformatic approaches for the prediction of the secondary structure of membrane proteins

Computational methods for predicting membrane topology can be very useful, and may provide information on TMS topology, helix-helix interactions, and functionally important regions of membrane proteins. Prediction methods have become more accurate over the years although several issues, concerning f.i. TMS length and position, tilted or broken TMSs, and reentrant loops, remain elusive. A wide variety of membrane topology prediction methods have been developed, using conceptually different strategies. In this section the development of prediction methods and a number of better known and representative predictors will be discussed.
The fact that TMSs of α-helical membrane proteins are on average more hydrophobic than loop regions, led to the development of early membrane topology prediction algorithms. The first method to predict the position of TMSs was based on hydropathy profile analysis alone [22]. The method classifies regions that are sufficiently hydrophobic as TMSs, using a sliding window of 19 residues to identify TMSs of 20-30 residues. Predictions improved significantly when the observation that positively charged amino acids are more abundant in cytoplasmic than in external loop regions ("the positive inside rule") was taken into account. In Toppred, regions that are of intermediate hydrophobicity can be classified either as TMS or loop region by optimizing the positive residues in cytoplasmic loops [34,35]. Additional improvement, in the form of statistical optimization, arrived with MEMSAT [36], which was the first method to apply an algorithm that guarantees a most likely global topology (with respect to the underlying residue preference values, the propensity scales). MEMSAT employs a set of statistical tables with residue preference values derived from well-characterized membrane protein data. Each table was associated with one of the five structural states (inside loop, inside helix end, middle helix, outside helix end, ouside loop, and showed definite bias toward certain amino acid species on the inside, middle and outside of the membrane. Optimal state assignments are found by a dynamic programming algorithm that computes scores for all possible topologies, in order to find the best possible match for the given data. A different prediction method that specialized modeling of various regions of a membrane protein, was based on a Hidden Markov Model (HMM) approach. HMMTOP [37] and TMHMM [2,38] were the first. The HMM layout of TMHMM is shown in

Fig. 3 (Fig. 1, [2]). Each box in the HMM layout corresponds to a submodel designed to model a specific region of a membrane protein. In order to model the length of the various regions, submodels contain several HMM states. The arrows show how the transitions between submodels can be made without disobeying the topological grammar of a helical membrane protein. The topological grammar provides a set of rules that defines how to form a valid membrane topology from topological elements. It involves allowed lengths of transmembrane regions and minimum lengths of inter-transmembrane loops, and state that cytoplasmic and non-cytoplasmic loops have to alternate. The TMHMM method combines hydrophobicity, charge bias, helix lengths, and topological grammar into one model.

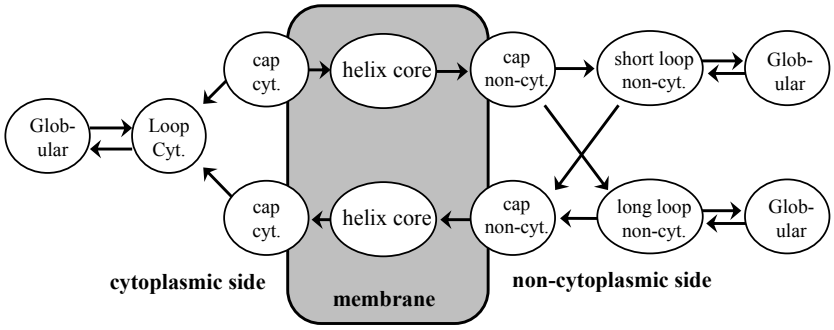


Fig. 3. The layout of a Hidden Markov Model. See text for further explanation. [Krogh et al. 2001]

Variations and/or combinations of the best approaches, e.g. in PRODIV-TMHMM [39], MEMSAT3 [40] and OCTOPUS [41], have led to additional improvements using f.i. evolutionary information. As PHDhtm and MEMSAT3 earlier, OCTOPUS [41] is based on residue preference scores derived from sequence profiles and artificial neural networks (ANNs), which are combined into a final topology. In particular, OCTOPUS includes regions that do not fully span the membrane (reentrant/membrane-dipping regions and helical hairpins) in its topological grammar.

TOPCONS (http://topcons.net/) combines an arbitrary number of topology predictions into one consensus prediction [42]. In this case, the reliability of the prediction is quantified based on the level of agreement between the underlying methods, both on the protein level and on the level of individual TM regions. A reported implementation of TOPCONS as a webserver was based on the individual topology predictors OCTOPUS [41], PRO-TMHMM and PRODIV-TMHMM

[39], SCAMPI-single and SCAMPI-multi [43]. Four of the topology predictors make use of multiple sequence information created using BLAST [28] to do a consensus prediction of transmembrane segments in the target protein, whereas the fifth method (SCAMPI-single) only requires the protein sequence. When TOPCONS was used to predict the membrane topology of CitS of *Klebsiella pneumoniae* a final membrane topological model with 11 TMSs was found, with a cytoplasmic N-terminus, a cytoplasmic central loop and a periplasmic C-terminus. At this level, the predicted topology is in good agreement with the experimentally determined topology [30,44,45,46], but does not recognize the two putative reentrant loops in the experimental model. As a matter of fact, the five individual topologies that were combined into the consensus prediction disagreed on two regions only, i.e. the (moderately) hydrophobic regions that correspond to the putative reentrant loops. PRO- and PRODIV-TMHMM both assigned TMSs, rather than reentrant loops, resulting in double inversion of the orientation with respect to the membrane. The predictors SCAMPI-single, SCAMPI-multi and OCTOPUS agreed 100% on the membrane topology of CitS, determining the final consensus prediction.

The overall prediction accuracy of a number of different methods was tested [41] using the following criteria: a correctly predicted topology includes predicting the correct number and location of all of the TMSs, as well as their orientations with respect to the membrane. The results are shown in table 2. On a dataset of 124 sequences, 94% of the topologies were correctly predicted by OCTOPUS, while PHDhtm scored lowest with 52%. Interestingly, the dataset contained more than 14 reentrant loops and/or membrane dips. Nearly all of these regions were recognized by OCTOPUS and MEMSAT3, in a manner that did not induce errors in other parts of the sequence. Apparently, these predictors are much less prone to mistake a reentrant loop region for a TMS than the other predictors.

Table 2.

Prediction accuracies and different types of prediction errors.

Method	Prediction accuracies (%)				
	Correct topologies[a]	Sequences with FNs[b]	Sequences with FPs[c]	Inverted topologies[d]	Dip or Reentrant FPs / all FPs[e]
OCTOPUS	94	4	2	1	100
MEMSAT3	87	7	6	1	56
PRODIV	81	5	11	5	83
PRO	77	12	10	3	92
TMHMM	60	17	7	17	88
PHDhtm	52	27	15	6	62

[a] Correct topologies - A topology prediction is correct if the number of predicted TM helices is accurate as well as their location with respect to the helices as defined in the experimental structures

[b] Sequences with False Negatives (FNs) and [c]False Positives (FPs) - The fraction of sequences for which a method misses, or overpredicts at least one TM region.

[d] Inverted topologies - The fraction of sequences for which a method predicts the TM regions correctly, but the orientation wrongly.

[e] Dip or Reentrant FPs / all FPs - The fraction of all false positive TM predictions that correspond to an actual reentrant or dip region. [Viklund et al. 2008]

4. Outline of this thesis

In Chapter 2, a membrane topology screening method is proposed by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method, TopScreen, was evaluated using 16 transporter proteins of known function from 4 different structural classes, and the experimental data was used to discriminate between membrane topology models from TMHMM, MemGen and a number of more recently developed predictors.

Class ST[3] of the MemGen classification contains 32 families of transporter proteins including the IT superfamily. In Chapter 3, secondary transporters from 19 different families in class ST[3] were evaluated by the TopScreen experimental topology screening method described in Chapter 2 to verify the structural classification by MemGen. It is concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other

families in class ST[3]. The core structure of the model consists of two homologous domains that contain 5 transmembrane segments each, with opposite orientations in the membrane. A reentrant loop is present in between the 4th and 5th segments in each domain.

Chapter 4 discusses the topological organization of the tartrate transporter TtdT from *Escherichia coli,* a member of the DASS family in structural class ST[3] of the MemGen classification, which was analyzed by in-frame translational fusions of progressively truncated forms of the *ttdT* gene and downstream reporter genes (*gfp* and *phoA*), and by cysteine accessibility studies. While the cysteine accessibility studies were inconclusive, the fusion data was consistent with the core structure of the ST[3] model. The L-tartrate:succinate exchange activity of the cloned TtdT transporter was verified, and the correlation between TtdT expression levels and transport activity was analyzed.

References

1. E. Wallin, and G. von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, *Protein Sci*. **7(4)** (1998) 1029-38.
2. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* **305** (2001) 567-580.
3. M.H. Saier, A functional-phylogenetic classification system for transmembrane solute transporters, *Microbiol. Mol. Rev*. **64** (2000) 354-411.
4. M.H. Saier Jr, M.R. Yen, K. Noto, D.G. Tamang and C. Elkan, The Transporter Classification Database: recent advances, Nucl. Acids Res. 37 (2009) D274-D2788.
5. S.S. Pao, I.T. Paulsen, M.H. Saier Jr., Major Facilitator Superfamily, *Microbiol. Mol. Biol. Rev*. **62(1)** (1998) 1–34.
6. D.L. Jack, I.T. Paulsen, M.H. Saier, The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations, *Microbiology* **146(8)** (2000) 1797-1814.
7. S. Prakash, G. Cooper, S. Singhi, M.H. Saier Jr., The ion transporter superfamily, *Biochim. Biophys. Acta.* **1618** (2003) 79-92.
8. Z. Jia, Z. Pei, D. Maiguel, C.J. Toomer, and P.A. Watkins, The fatty acid transport protein (FATP) family: very long chain acyl-CoA synthetases or solute carriers?, *J. Mol. Neurosci* **33** (2007) 25-31.
9. F. Katzen, and J. Beckwith, Transmembrane electron transfer by the membrane protein DsbD occurs via a disulfide bond cascade, *Cell* **103** (2000) 769-779.
10. R. Krupp, C. Chan, D. Missiakas, DsbD-catalyzed transport of electrons across the membrane of *Escherichia coli*, *J. Biol. Chem*. **276** (2001) 3696-3701.
11. S.H. Cho, and J. Beckwith, Mutations of the Membrane-Bound Disulfide Reductase DsbD That Block Electron Transfer Steps from Cytoplasm to Periplasm in *Escherichia coli*, *J. Bacteriol*. **188** (2006) 5066-5076.
12. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*, *Science* **301** (2003) 610-615.
13. Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*, *Science* **301** (2003) 616-620.

14. J. S. Lolkema, and D. J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, *Mol. Membr. Biol*. **15** (1998) 33-42.
15. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters, *Nature* **437** (2005) 215-223.
16. S. Faham, A. Watanabe, G. Mercado Bessemer, D. Cascio, A. Szpecht, B. A. Hirayama, E.M. Wright, J. Abramson, The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport, *Science* **321** (2008) 810-814.
17. S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E.P. Carpenter, N.G. Rutherford, J.M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S.G. Patching, R.J. Hope, H.T. Norbertczak, P.C. Roach, S. Iwata, P.J. Henderson, A.D. Cameron, Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter, *Science* **322** (2008) 709-713.
18. S. Ressl, A.C. Terwisscha van Scheltinga, C. Vonrhein, V. Ott, C Ziegler, Molecular basis of transport and regulation in the Na+/betaine symporter BetP, *Nature* **458** (2009) 47-52.
19. Y. Fang, H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, C. Miller, Structure of a prokaryotic virtual proton pump at 3.2 A resolution, *Nature* **460** (2009) 1040-1043.
20. Xiang Gao, Feiran Lu, Lijun Zhou, Shangyu Dang, Linfeng Sun, Xiaochun Li, Jiawei Wang, and Yigong Shi, Structure and Mechanism of an Amino Acid Antiporter, *Science* **324** (2009) 1565-68.
21. J.S. Lolkema, and D.J. Slotboom, The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-Leucine transporters, *Mol. Membr. Biol.* **25** (2008) 567-570.
22. J. Kyte, and R.F Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* **157(1)** (1982) 105-32.
23. J.S. Lolkema, and D.J. Slotboom, Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins, *FEMS Microbiol. Rev*. **22** (1998) 305-322.
24. J. S. Lolkema, and D. J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, *Mol. Membr. Biol*. **15** (1998) 33-42.

25. S.B. Needleman and C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins*, J. Mol. Biol.* **48** (1970) 443-453.
26. J.S. Lolkema, and D.J. Slotboom, Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis, *J. Mol. Biol*. **327** (2003) 901-909.
27. J.S. Lolkema, and D.J. Slotboom, Sequence and Hydropathy Profile Analysis of two classes of secondary transporters, *Mol. Membr. Biol* **22** (2005) 177-189.
28. S.F. Altschul, T.L. Madded, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res*. **25** (1997) 3389-3402.
29. J.S. Lolkema, I. Sobczak, D.J Slotboom, Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops, FEBS Journal **272** (2005) 2334–2344.
30. I. Sobczak and J. S. Lolkema, The 2-hydroxycarboxylate transporter (2HCT) family: Physiology, Structure and Mechanism, *Microbiol. Mol. Biol. Rev*. **69** (2005) 665-695.
31. J.S. Lolkema, Domain structure and pore-loops in the 2-hydroxycarboxylate transporter family, *J. Mol. Microbiol. Biotechnol*. **11** (2006) 318-325.
32. D. Yernool, O. Boudker, Y. Jin, and E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, *Nature* **431** (2004) 811-818.
33. A. J. Dobrowolski, I. Sobczak, and J. S. Lolkema, Experimental Validation of Membrane Topology Prediction by Hydropathy Profile Alignment: Membrane Topology of the Na^+-Glutamate Transporter of *Escherichia coli*, *Biochemistry* **46** (2007) 2326-2332.
34. G. von Heijne, Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule, *J. Mol. Biol*. **225(2)** (1992) 487-494.
35. M.G. Claros, and G. Von Heijne, TopPred II: an improved software for membrane protein structure predictions, *Comput. Appl. Biosci.* **10(60)** (1994) 685-686.
36. D.T. Jones, W.R. Taylor, J.M. Thornton, A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology, *Biochemistry* **33(10)** (1994) 3038–3049.

37. G.E. Tusnády, and I. Simon, Principles governing amino acid composition of integral membrane proteins: application to topology prediction, *J. Mol. Biol*. **283(2)** (1998) 489-506.
38. E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6** (1998) 175-82.
39. H. Viklund, and A. Elofsson, Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information, *Protein Sci.* **13(7)** (2004) 1908-17.
40. D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics* **23** (2007) 538-544.
41. H. Viklund and A. Elofsson, Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar, *Bioinformatics* **24** (2008) 1662-1668.
42. A. Bernsel, H. Viklund, A. Hennerdal and A. Elofsson, TOPCONS: consensus prediction of membrane protein topology, *Nucleic Acids Research* **37** (2009) Web Server issue W465-68.
43. A. Bernsel, H.Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elofsson, Prediction of membrane-protein topology from first principles, *Proc. Natl. Acad. Sci. U.S.A*. **105(20)** (2008) 7177-81.
44. M. van Geest, and J. S. Lolkema, Membrane topology of the sodium ion-dependent citrate carrier of *Klebsiella pneumoniae*. Evidence for a new structural class of secondary transporters, *J. Biol. Chem*. **271** (1996) 25582-25589.
45. M. van Geest and J. S. Lolkema, Membrane topology of the Na(+)/citrate transporter CitS of *Klebsiella pneumoniae* by insertion mutagenesis, *Biochim. Biophys. Acta* **1466** (2000) 328-338.
46. I. Sobczak and J. S. Lolkema, Accessibility of cysteine residues in a cytoplasmic loop of CitS of *Klebsiella pneumoniae* is controlled by the catalytic state of the transporter, *Biochemistry* **42** (2003) 9789-9796.

Chapter 2

Rapid Screening of Membrane Topology of Secondary Transport Proteins

Ramon ter Horst and Juke S. Lolkema

Biochim. Biophys. Acta. **1798(3)** (2010) 672-80

Limited experimental data may be very useful to discriminate between membrane topology models of membrane proteins derived from different methods. A membrane topology screening method is proposed by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method involves amplification of the encoding genes or gene fragments by PCR, rapid cloning in dedicated vectors by ligation independent cloning, and determination of the cellular disposition of the three sites using conventional techniques. The N-terminus was determined by labeling with a fluorescent probe, the central position and the C-terminus by the reporter fusion technique using alkaline phosphatase (PhoA) and green fluorescence protein (GFP) as reporters.

The method was evaluated using 16 transporter proteins of known function from 4 different structural classes. For 13 proteins a complete set of 3 localizations was obtained. The experimental data was used to discriminate between membrane topology models predicted by TMHMM, a widely used predictor using the amino acid sequence as input and by MemGen, that uses hydropathy profile alignment and known 3D structures or existing models. It follows that in those cases where the models from the two methods were similar, the models were consistent with the experimental data. In those cases where the models differed, the MemGen model agreed with the experimental data. Three more recent predictors, MEMSAT3, OCTOPUS and TOPCONS showed a significantly higher consistency with the experimental data than observed with TMHMM.

1. Introduction

Secondary transporters catalyze the translocation of substrates across membranes driven by (electro) chemical gradients of substrates and co-ions. They are universal to all biological cells and involved in many biological processes. They are integral membrane proteins that in most cases are encoded by a single gene. In spite of their simple architecture, their phylogenetic diversity is enormous as evidenced by the approximately 100 gene (super)families in the Electrochemical Potential-driven transporters section of the Transport Classification system [1]. Most likely, the genetic diversity is at least in part a consequence of divergent evolution and, consequently, many of the different families may represent a similar fold and translocation mechanism. Support for this view is provided by the recently reported high resolution structures of the Na^+-leucine transporter LeuT [2], the Na^+-galactose transporter vSGLT [3], the Na^+-benzyl-hydantoin transporter Mhp1 [4], the Na^+-betain transporter BetP [5] and, very recently, the arginine/agmatine exchanger AdiC [6,7] that all show the same core structure. The five transporters are from different gene families and no significant sequence similarity could be identified between the transporters. Structural similarity between transporters from these families was predicted before by the MemGen method that is based on hydropathy profile analysis of amino acid sequences of membrane proteins [8,9]. MemGen groups families of transporters with the same global fold in structural classes and the 5 transporters LeuT, vSGLT, Mhp1, BetP and AdiC are all found in structural class ST[2] [10]. Classification of all transporter families, and, more in general, all membrane proteins, in structural classes by a computational technique like MemGen would be a major step forward in membrane protein research.

At a much lower level of resolution, a membrane topology model describing the number of transmembrane segments (TMS) and the orientation of the protein in the membrane, is an essential first step in any study of the structure/function relationships of a membrane protein. Computational analysis of the primary structure is especially amenable for membrane topology prediction due to the back and forward folding of the polypeptide chain through the hydrophobic phospholipid bilayer, and many methods have been developed and are used extensively. It has been argued that the prediction methods may be significantly improved by constraining the prediction algorithm with limited experimental data [11]. The cellular location of the C-terminus of a significant part of the membrane proteome of *Escherichia coli* [12,13] and *Saccharomyces cerevisiae* [14] was determined experimentally by fusing the reporters alkaline phosphatase

(PhoA) and green fluorescent protein (GFP) to the proteins. Subsequently, this data was used to produce improved topology models by constraining the location of the C-terminus of the proteins. Using homology, the experimentally constrained topology models for the *E. coli* and *S. cerevisiae* proteome could be assigned to tens of thousands other membrane proteins in the public databases [14,15].

In a different approach, limited experimental data may also be used to discriminate between topology models produced by different methods. MemGen structural class ST[3] contains 33 families of bacterial secondary transporters for organic and inorganic anions and Na^+/H^+ antiporters [16]. MemGen is not a secondary structure prediction method, but since the fold of the proteins in the different families in one class is identical, membrane topology for the different families may be predicted by hydropathy profile alignment using a known model of one of the families [16,17]. Topology models for the different families in ST[3] were predicted on the basis of the experimentally determined topology of CitS of *Klebsiella pneumoniae*, a member of the 2HCT family in ST[3] [18,19]. For many, if not all families the MemGen models differ from the models produced by TMHMM [20], a widely used secondary structure predictor that produces topology models using the amino acid sequence as input. The global topology analysis of the *E. coli* proteome [12] revealed the C-terminal location of 19 proteins in structural class ST[3] that were distributed over 7 different families. The MemGen models for the proteins showed a C-terminal location that was in agreement with the experimentally determined location in all cases, while TMHMM predicted the C-terminus correctly only in roughly half of the cases [21]. Remarkably, constraining the C-terminus of the proteins in the TMHMM prediction [11] resulted in consensus between the MemGen and TMHMM methods for only 3 out of 19 proteins.

It follows that the experimental localization of a single site in a protein sequence is not sufficient to discriminate between different topology models. Here we present and explore a method in which the localization of three sites in the sequence are determined experimentally, the N- and C-terminus and a site in the middle of the protein. Where a single site may just detect an erroneously predicted orientation of the protein in the membrane, multiple sites are more likely to detect missing or extra transmembrane segments in the model. Moreover, selection of a site in the middle of the sequence acknowledges that many membrane proteins are two-domain proteins. A missed transmembrane segment in the first domain might be compensated for by a similar error in the second domain, which would not be identified by the location of the N- and C-termini. The

method was tested on a set of 16 secondary transporters from 4 different structural classes defined by MemGen including class ST[3], thereby representing 4 different structures. Experimental results were used to evaluate the membrane topology models predicted by MemGen and TMHMM and three more recently presented predictors, TOPCONS, MEMSAT3 and OCTOPUS.

2. Materials and Methods

2.1. Materials

Phusion DNA polymerase was obtained from Finnzymes (Espoo, Finland). T4 ligase was obtained from New England Biolabs (Frankfurt am Main, Germany). All other enzymes were obtained from Fermentas (Burlington, Canada). Mutagenic oligonucleotides were obtained from Biolegio (Nijmegen, The Netherlands), or from Operon (Ebersberg, Germany) for ligation independent cloning. *p*-Nitrophenyl phosphate (pNPP) was obtained from Sigma (Zwijndrecht, The Netherlands), fluorescein maleimide (FM) from Invitrogen (Carslbad, United States)

2.2. Bacterial strains and growth conditions

Escherichia coli strain SF100 (*recA Δlac ΔompT*) [22] harboring the indicated pLIC vector (see below), was routinely grown in Luria Broth medium at 37 ^{o}C, with ampicillin added at a final concentration of 50 μg/ml. Overnight cultures were diluted 30-fold in 3 mL of fresh medium and when the optical density measured at 660 nm (OD_{660}) reached a value between 0.6 and 0.8, arabinose was added at a final concentration of 0.002-0.05 % (w/v) to induce protein production from the plasmids. Following growth for another 1.5-2 h, cells were harvested by centrifugation in a table top centrifuge operated at 4 ^{o}C. Cells were resuspended in the indicated buffer and kept on ice until use.

2.3. Construction of LIC vectors and Ligation Independent Cloning.

All genetic manipulations were done using standard techniques. A *KpnI/XbaI* fragment of pPHO7 [23] containing the *phoA* gene encoding alkaline phosphatase without its signal sequence was ligated downstream of the arabinose promoter in the commercial pBAD24 vector (Invitrogen) [24] using the same two restriction enzymes. The resulting vector was digested with *NcoI* and *KpnI* and a synthetic double stranded piece of DNA (the LIC cassette) with *NcoI* and *KpnI* compatible overhang at the 5' and 3' ends, respectively, was inserted. The double stranded DNA fragment was prepared by hybridizing two oligonucleotides. The nucleotide sequence of the sense strand was 5'-C ATG GGT *CAT CAT CAC CAT CAC CA**T*** **TTA AA**T AGT GGT GTG GTA C- 3' in which the initiation codon was underlined, the sequence coding for a His_6-tag set in italics and a *SwaI* restriction site in bold. The resulting vector, termed pLIC1 was used to construct alkaline phosphatase fusion proteins (see below). Vector pLIC2, which is used for GFP fusions, was constructed by replacing the *phoA* gene in pLIC1 by the *gfp* gene encoding green fluorescence protein. A *KpnI/XbaI* fragment of pLIC1 was replaced by a fragment taken from the vector pBADcLIC_GFP [25] digested with *KpnI* and *XbaI.* Similarly, the *phoA* fragment was replaced by a double stop codon resulting in vector pLIC3 which is used to produce His-tagged proteins. The second codon of the cassette in pLIC3, GGT (Gly) was mutated into TGT (Cys) by site directed mutagenesis rendering vector pLIC4 which is used for N-terminal localizations.

Ligation Independent Cloning (LIC) was done essentially as described [25]. The LIC vectors pLIC1-4 were linearized by *Swa*I digestion. Transporter genes were amplified using forward and backward primers containing 5' flanking regions corresponding to the nucleotide sequences upstream and downstream of the *Swa*I site in the LIC cassette as follows: 5'-CATGGGTCATCATCACCATCACCATTG.....-3' and 5'-TACCACACCACTATTTTG.....-3', respectively. For each individual gene one forward and two reverse primers were used to amplify full length and half length transporter genes (see also Figure 1). Single-stranded overhangs of the PCR products and vectors were generated using T4 DNA polymerase in the presence of dGTP (vector) and dCTP (PCR product). The complementary overhangs of PCR product and vector annealed upon mixing, after which the resulting heteroduplexes were transformed to *E. coli.*

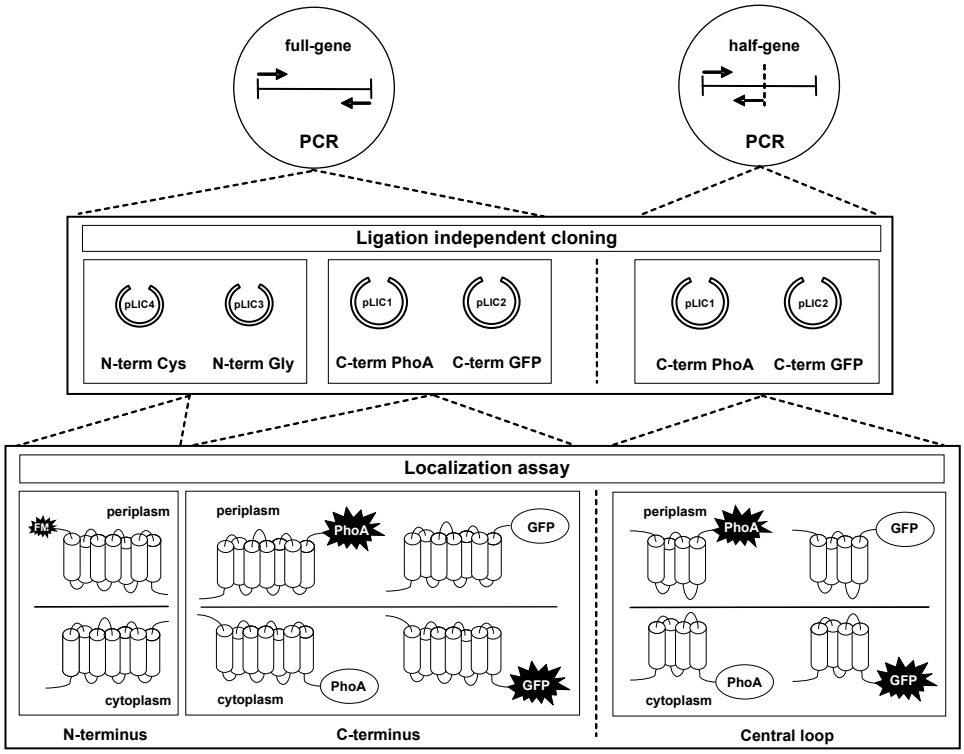


Figure 1. Schematic representation of the membrane topology screening method. The method consists of 3 steps: (i) amplifying of the full-length gene and half-gene by PCR, (ii) cloning into a set of dedicated vectors by ligation independent cloning, and (iii) the localization assays. The dedicated vectors introduce a Cys (pLIC4) or Gly (pLIC3) residue at the N-terminus, or fuse PhoA (pLIC1) or GFP (pLIC2) at the C-terminus of the protein encoded by the insert. The N-terminal localization of the full-length protein is determined by the accessibility of the introduced cysteine residue at the external face of the membrane using FM, a membrane impermeable fluorescent sulfhydryl reagent. The C-terminal localizations of the full length and half proteins are reported by PhoA (periplasmic) and GFP (cytoplasmic) activity. For further details, see the text.

2.4. GFP and PhoA assays

2.4.1. PhoA assay

Cells from 0.5 ml of culture of *E. coli* SF100 cells harboring a pLIC1 vector carrying the indicated insert were washed once, and subsequently, resuspended in 1 M Tris-HCl pH 8.0 buffer after which the OD_{660} was measured. Following incubation for 5 min at 37 ^{o}C, 200 µL of a 1.5 mM *p*-nitrophenyl phosphate (pNPP) solution was added and the suspension was incubated until a yellow color developed. Then, the cells were spun down and the absorption was read at 420 and

550 nm using a Hitachi U-1100 spectrophotometer. The phosphatase activity was calculated in Miller units [26].

2.4.2. GFP assay.

Cells of *E. coli* SF100 harboring a pLIC2 vector carrying the indicated insert were washed once and, subsequently, resuspended in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 15 mM EDTA to a final OD_{660} of 0.2. Following incubation for 30 min at room temperature, 150 µL cell suspension was transferred into a precision cell (Hellma, Quartz SUPRSIL) and GFP fluorescence emission intensity at 508 nm was measured at an excitation wavelength of 470 nm using an AMINCO Bowman Series 2 Luminescence Spectrometer. Background signals caused by light scattering were estimated from cells harboring the pLIC3 vector without insert [27].

Evaluation of data. Mean values and standard deviations were calculated from at least three independent measurements. PhoA and GFP activities of the cells were normalized by the mean PhoA activity of all positive PhoA fusions (>100 Miller units) and all positive GFP fusions (> 0.5 emission units), respectively. The logarithm of the ratio of the normalized PhoA and GFP activities was calculated for each full-length and half protein to obtain a measure for the cellular localization of the fusion point. Values >2.5 or <-2.5 were arbitrarily set to 2.5 and -2.5, respectively.

2.5. Labeling studies

E. coli SF100 cells harboring pLIC4 or pLIC3 vectors carrying the indicated insert were washed once and, subsequently, resuspended in ice-cold 50 mM potassium phosphate buffer pH 7.0. Cells from 200 ml of culture were resuspended in 1 ml of buffer. A 5 mM solution of fluorescein 5-maleimide (FM) was freshly prepared by diluting a 50 mM solution in DMSO with 50 mM potassium phosphate buffer pH 7.0. FM was added to the cell suspension to a final concentration of 100 µM. After incubation for 20 min at room temperature, excess FM was quenched by adding 2 mM dithiothreitol (DTT). Cells were washed three times with 50 mM potassium phosphate buffer pH 7.0, and disrupted by sonication. Following removal of debris, membranes were collected from the supernatant by centrifugation for 25 min at 80,000 rpm in a Beckman TLA100.2 rotor at 4 ^{o}C. The pellet was solubilized in 1 mL 50 mM potassium phosphate buffer

pH 8.0 containing 400 mM NaCl, 20% glycerol, 10 mM imidazole, 1mM DTT and 2% Triton X-100 and left on ice for 30 min after which undissolved material was removed by centrifugation at 90,000 rpm for 25 min at 4 ^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (25 µL bed volume) equilibrated in 50 mM potassium phosphate buffer pH 8.0 containing 600 mM NaCl, 20% glycerol, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100, and incubated overnight at 4^{o}C under continuous shaking. The resin was pelleted by brief centrifugation in a table top centrifuge and the supernatant was removed. The resin was washed with 0.5 mL of equilibration buffer containing 300 mM NaCl and 20-40 mM imidazole. The protein was eluted with 25 µL of the same buffer, but at a pH of 7.0 and containing 200 mM imidazole. Subsequently, samples were loaded onto a 12% sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE). Following electrophoresis, in-gel fluorescence was recorded using a Fujifilm LAS-4000 luminescent image analyzer, and the gel was stained with Coomassie Brilliant Blue (CBB).

3. Results

3.1. Strategy

To discriminate between different topology models of membrane proteins, the cellular disposition of three positions in each of the polypeptide chains was determined by a combination of conventional techniques. The locations of the C-termini of full-length proteins and of half-proteins, truncated at a central loop, were determined using reporter gene fusion techniques. Green fluorescent protein (GFP) and alkaline phosphatase (PhoA), devoid of its signal sequence, were used as positive reporters of a cytoplasmic and periplasmic localization, respectively. GFP folds into an active conformation when localized in the cytoplasm, but not in the periplasm [28]. In contrast, PhoA matures into the active state only when residing in the periplasm [29]. The third position, the N-terminus, was determined by the accessibility of an engineered cysteine residue using a membrane impermeable fluorescent sulfhydryl reagent.

The procedure as outlined in Figure 1 was set up using rapid, ligation independent cloning (LIC) [25] and a set of dedicated cloning vectors (see Methods) to allow for the screening of many proteins in high throughput. A single PCR product encoding the full-length protein was cloned into a set of four cloning vectors resulting in the fusions with PhoA and GFP at the C-terminus of

the protein and the introduction of a Cys and Gly residue at position 2 at the N-terminus. The construct with the Gly residue was used as a control in the labeling studies. The Cys and Gly residues were placed upstream of a His_6-tag that was present in all constructs to allow for maximum accessibility of the Cys residue. A second PCR product encoding the N-terminal half of the protein (half-protein) was cloned into the two vectors resulting in the PhoA and GFP fusion proteins. Whenever possible, the protein was truncated in between the (putative) N-terminal and C-terminal domain.

Localization assays were done with resting cells using standard assays for alkaline phosphatase activity and fluorescence measurements for GFP content. Cells expressing the appropriate construct were treated with the fluorescent membrane-impermeable sulfhydryl reagent, fluorescein maleimide (FM) followed by processing of the cells to determine N-terminal location by the labeling efficiency.

3.2. Selection of transporter proteins and topology models

A set of 16 secondary transport proteins of known function was selected to evaluate the procedure outlined in Figure 1. All transporters were from bacterial origin, originating from the phyla proteobacteria and firmicutes (Table 1). They were selected from 4 different structural classes in the MemGen classification and, therefore, putatively represent four different structures. The transporters from one structural class were mostly selected from different gene families (Transport Classification system; TC system [1]), thereby warranting high amino acid sequence diversity. In general, no significant sequence similarity can be detected between transporters from different families in one structural class. Secondary structure predictor TMHMM [20] predicts a topology model for each single sequence. In contrast, MemGen needs a template structure to model membrane topologies of other proteins in the same structural class by hydropathy profile alignment [8,9,10,16,17].

Table 1.

Transporter proteins used in this study

MemGen class	Transport protein	TC superfamily/family	Organism	Lineage[1]	Function	Half-protein[2]	Reference
ST[1]	KgtP	MFS/MHS	*Escherichia coli*	B-P-c	α-ketoglutarate:H^+ symporter	S222	42
	LmrP	MFS/DHA	*Lactococcus lactis*	B-F-l	multidrug:H^+ antiporter	P193	43
	MelB	MFS/GPH	*Escherichia coli*	B-P-c	melibiose:Na^+ symporter	S203	44
ST[2]	AguD	APC/APA	*Lactobacillus brevis*	B-F-l	agmatine/putrescine exchanger	S180/M216[c]	45
	HdcP	APC/GGA	*Lactobacillus hilgardi*	B-F-l	histidine/histamine exchanger	A195/A238[c]	46
	TyrP	HAAAP	*Lactobacillus brevis*	B-F-l	tyrosine/tyramine exchanger	L183/P222[c]	47
ST[3]	CimH	2HCT	*Bacillus subtilis*	B-F-b	*L*-malate/citrate:H^+ symporter	E261	48
	CitH	IT/CitMHS	*Bacillus subtilis*	B-F-b	Ca^{2+}-citrate:H^+ symporter	R202	49
	CitM	IT/CitMHS	*Bacillus subtilis*	B-F-b	Mg^{2+}-Citrate:H^+ symporter	E208	49
	CitP	2HCT	*Leuconostoc mesenteroides*	B-F-l	citrate/lactate exchanger	V269	50
	CitS	2HCT	*Klebsiella pneumonia*	B-P-c	citrate:Na^+ symporter	S251	50
	GltS	ESS	*Escherichia coli*	B-P-c	glutamate:Na^+ symporter	Q197	51
	MleP	2HCT	*Lactococcus lactis*	B-P-c	malate/lactate exchanger	E240	50
ST[4]	DctA	DAACS	*Escherichia coli*	B-P-c	C_4-dicarboxylate transporter	S271	52
	GltT	DAACS	*Bacillus stearothermophilus*	B-F-b	glutamate:H^+ symporter	T272	52
	SSTT	DAACS	*Escherichia coli*	B-P-c	serine/threonine:Na^+ symporter	A269	52

[a] B-P-c: Bacteria, Proteobacteria, gamma subdivision; B-F-l: Bacteria, Firmicutes, lactobacillales; B-F-b: Bacteria, Firmicutes, bacillales.

[b] Last residue of wild-type sequence in N-terminal half-protein.

[c] Five TMS/six TMS half-proteins, respectively.

The three transporters selected from MemGen structural class ST[1], KgtP, LmrP, and MelB are from different gene families in the same superfamily: the Major Facilitator Superfamily (MFS) in the TC system [1]. The MemGen topology model of the ST[1] transporters is based on the high resolution structures of the LacY and GlpT transporters [30,31] and shows two (homologous) domains containing 6 TMSs each with both termini in the cytoplasm (Figure 2). Two of the three exchangers selected from MemGen structural class ST[2] are from different families in the APC superfamily and a third (TyrP) from yet another family, HAAAP. High resolution structures showing the same fold of the core of the proteins are available for transporters from five unrelated families in ST[2] (see Introduction). Hydropathy profile alignment reveals topology models for the selected transporters with two (homologous) domains containing 5 TMSs each, oriented in opposite directions in the membrane (inverted topology). The transporters are predicted to have two or three additional TMSs at the C-terminus (Figure 2). Seven transporters were selected from three different families in structural class ST[3]. The CitMHS family is found in the IT (Ion Transporters) superfamily. No high resolution structure is available for a transporter in class ST[3]. The topology model is based on extensive studies of the membrane topology of CitS of *Klebsiella pneumoniae* (2HCT family) and GltS of *Escherichia coli* (ESS family) that both were included in the set as references [17,18,19]. The model reveals two (homologous) domains containing 5 TMSs each resulting in inverted topology. In between the fourth and fifth TMS in each domain the loop region folds back in between the TMSs (reentrant or pore-loop). The 2HCT family proteins have an additional TMS at the N-terminus. Three transporters were selected from ST[4]. ST[4] consists of a single family (DAACS). GltT shares 38% and 19% sequence identities with DctA and SstT, respectively, while the identity between the latter two is 19%. The high-resolution structure of the Na^+-aspartate transporter Glt_{Ph} [32], was used to determine the membrane topology by MemGen. The models reveal 8 TMS with two reentrant loops in the C-terminal half of the proteins.

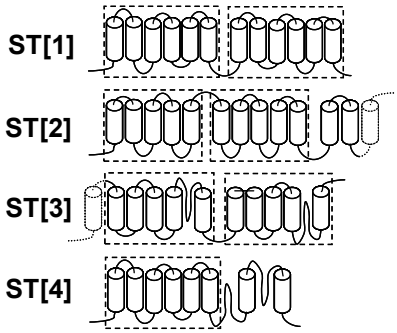


Figure 2. Membrane topology models of structural classes ST[1] to [4] by MemGen. See text for explanation. Boxed parts represent domains. Transmembrane segments indicated by thin lines are not present in all of the proteins in one class listed in Table 1.

3.3. C-terminal location of full-length and half-proteins

The alkaline phosphatase activity of cells expressing PhoA fusions of the full-length and half-proteins of each of the transporters in Table 1 was plotted against the GFP fluorescence of cells expressing the corresponding GFP fusion (Figure 3A,B). The data points formed two well-separated groups falling along the two axes. The two groups correspond to a periplasmic and cytoplasmic localization of the fusion point. In general, there was a good correlation between the two reporters: proteins resulting in a high PhoA activity showed a low GFP activity and *vice versa*. The 'negative' reporter signal observed for alkaline phosphatase was clearly lower than observed for GFP, suggesting a significant background of fluorescence intensity in the GFP assay. The dashed lines tentatively limit the regions in the graph where the two groups of points are found. The results for two full-length transporters were inconclusive. Signals obtained for GltS of *Escherichia coli* were insignificant (Figure 3A, arrow a), suggesting lack of expression of the fusion proteins, and MleP of *Lactococcus lactis* (arrow b) showed a relatively small, but significant positive signal with both reporters. Both these transporter are from structural class ST[3].

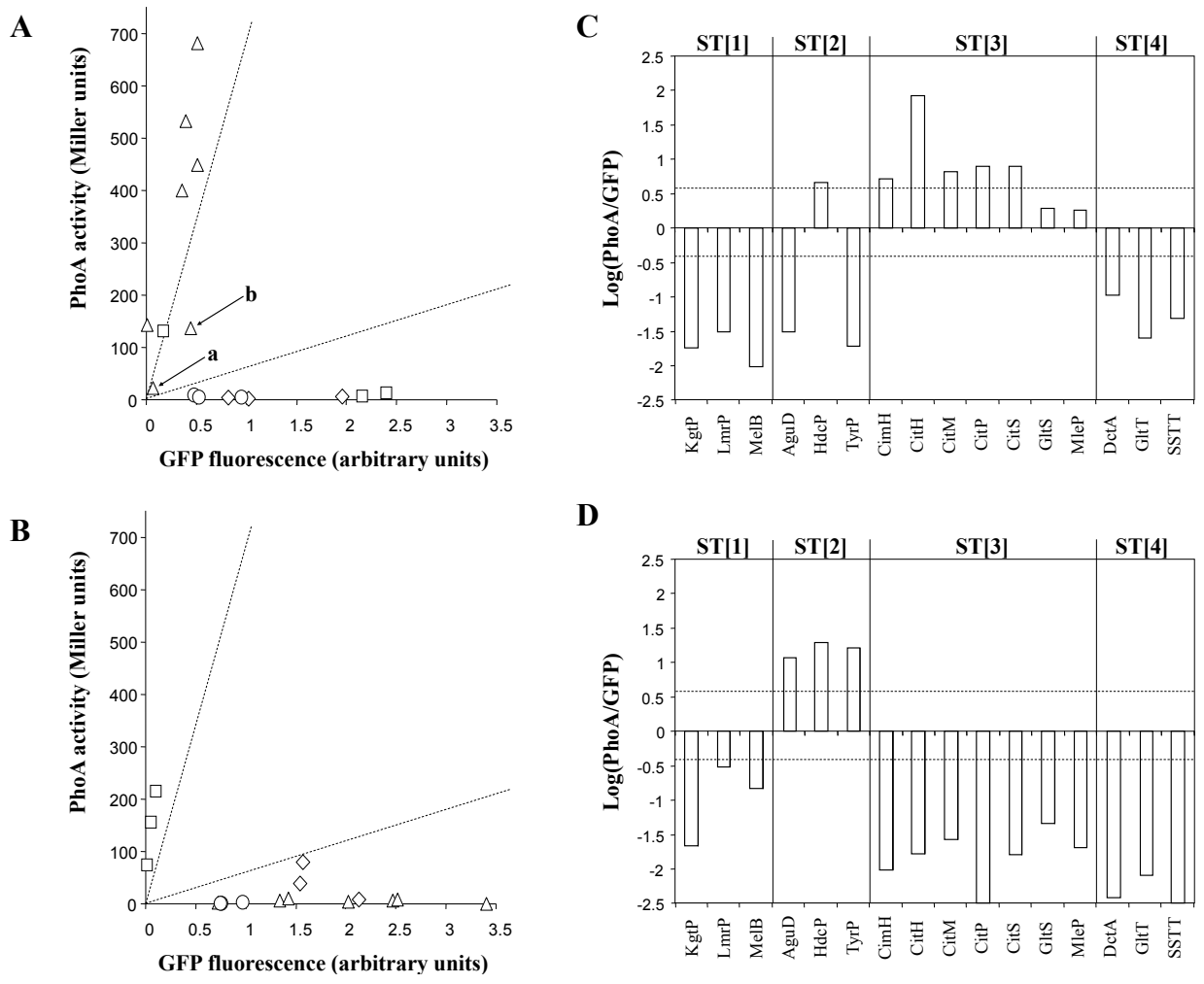


Figure 3. C-terminal localization of full-length and half proteins. (A,B) PhoA activity plotted against the GFP fluorescence intensities for full length proteins (A) and half-proteins (B) of the transporters listed in Table 1. ○, ST[1] proteins, □, ST[2] proteins Δ, ST[3] proteins, ◊, ST[4] proteins. Dotted lines indicate the limits for a periplasmic and cytoplasmic localization. (C,D) Logarithm of the ratio of the normalized PhoA and GFP signals for each of the full-length proteins (C) and half proteins (D). Threshold values for positive localizations were indicated by the dotted lines, which correspond to the dotted lines in panels A and B.

The range of values of positive PhoA activities and GFP fluorescence intensities was significant, and, mostly likely, represents the different levels of expression of the fusion proteins. Under the assumption that expression levels are mainly determined by the membrane protein and less so by the reporter protein, expression levels may be eliminated from the data by taking the ratio of the

PhoA activity and the GFP fluorescence intensity after normalization of the two signals (see Methods, section 2.4). Figures 3C and D show the logarithms of the normalized ratio of the PhoA and GFP signals for each of the full-length proteins (C) and half proteins (D). Then, a positive bar corresponds to a periplasmic localization, a negative bar to a cytoplasmic localization. The length of the bars indicates the significance of the localization and would be independent of the expression levels. Threshold values for positive localizations were indicated by the dashed lines that correspond to the dashed lines in Figures 3A and B.

The C-termini of the full-length transporters of structural classes ST1] and ST[4] are all found in the cytoplasm, while ST[3] transporters have their C-termini in the periplasm. As mentioned above, the results for GltS and MleP in ST[3] were inconclusive. In ST[2], the AguD and TyrP proteins end in the cytoplasm, while HdcP ends in the periplasm. The C-termini of the half proteins of ST[1], ST[3] and ST[4] are all found in the cytoplasm, while the N-terminal half-proteins of the ST[2] transporters end in the periplasm.

3.4. Domain structure of ST[2] transporters

TMHMM predicts 12 TMSs for most of the APC superfamily transporters in structural class ST[2], while MemGen models the same proteins using the available high resolution structures of transporters in ST[2] which results in two domains of 5 TMSs each plus two or three additional TMSs (see above). The half-proteins of the ST[2] transporters in the studies above were based on the latter prediction i.e. the half-proteins consisted of the predicted first domains (5 TMSs). On the other hand, the TMHMM topology might suggest a two times six TMSs structure like the transporters of the MFS in class ST[1]. Since the choice of the truncation point may affect the stability of the fusion protein and even erroneously report the localization [33], the fusions of the half-proteins of the three ST[2] transporters were also constructed based on this assumption, i.e. consisting of the first 6 TMSs. The results are presented in Figure 4. In case of the AguD and TyrP half-proteins, the C-terminus of the 6^{th} TMS is in the cytoplasm which is consistent with the result from the 5 TMS half-proteins. However, the results for the HdcP transporter were found to be inconclusive, both a high PhoA and GFP signal were observed. It follows that the 5 TMS half-proteins give a more consistent result than the 6 TMS half-proteins.

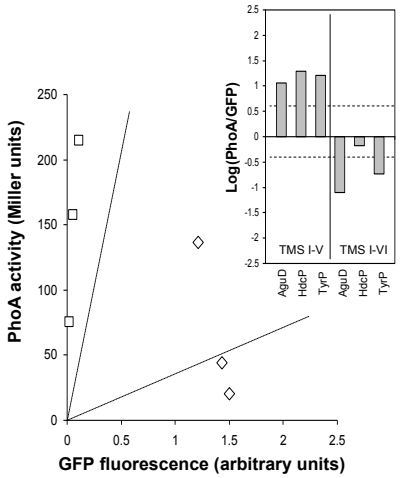


Figure 4. C-terminal localization of half-proteins in structural class ST[2]. PhoA activity plotted against the GFP fluorescence intensities for ST[2] proteins truncated after TMS V (□) and TMS VI (◊). Dotted lines correspond to the corresponding lines in Figures 3A,B. Insert. Logarithm of the ratio of the normalized PhoA and GFP signals for each of the fusion proteins.

3.5. N-terminal location of full-length proteins

For each of the transporters in Table 1 two constructs were made to determine the location of the N-terminus, one variant with a Cys and one with a Gly residue at position 2, immediately upstream of the His-tag. The cytoplasmic localization of the N-terminus of CitS of *Klebsiella pneumoniae* has been well documented [34,35,36]. Treatment of resting cells expressing the Gly variant of CitS with 0.1, 0.33 and 1.5 mM of fluorescence maleimide for 20 minutes showed significant labeling of the protein only at the highest FM concentration (Figure 5A). Apparently, the 5 endogenous Cys residues of CitS were only poorly accessible for FM added to the extra-cellular medium. The same experiment with the Cys variant showed dramatically more labeling at the highest concentration of FM indicating that under these conditions the introduced Cys residue in the cytoplasmic N-terminus was accessible to the reagent. It also shows that successful labeling of a single Cys residue results in a much stronger signal than obtained with Coomassie Brilliant Blue staining of the protein. Labeling of the Cys variant was considerably less at a concentration of 0.33 mM FM and could only be marginally detected at a concentration of 0.1 mM. It follows that treatment of cells with 0.1 mM of FM for 20 minutes does not result in significant labeling of Cys residues located at the cytoplasmic face of the membrane. The N-

terminus of GltS of *Escherichia coli* is located in the periplasm [17]. Treatment of cells expressing the Gly variant with 0.1 mM FM for 20 minutes does not reveal any of the endogenous Cys residues in the fluorescence image of the gel (Figure 5B). In contrast to what was observed with CitS, the Cys variant was clearly labeled after the same treatment, showing that the introduced Cys residue was readily accessible for the reagent, consistent with a periplasmic location. Treatment of cells expressing the two variants of the transporters with 0.1 mM FM for 20 minutes and at room temperature was used to discriminate between a cytoplasmic and periplasmic location of the N-terminus of the proteins.

The CitM and CitH transporters of *B. subtilis* in structural class ST[3] gave the same results as observed for GltS. No labeling was observed with the Gly variant, while the Cys variant showed labeling (Figure 5B). It follows that these transporters expose their N-termini at the exterior of the cell surface. The amount of CitH protein obtained after the partial purification was very small indicating low expression, which is consistent with previous observations [37,38]. In addition, the more sensitive fluorescent signal shows that the CitH protein runs as a diffuse band on SDS-PAGE.

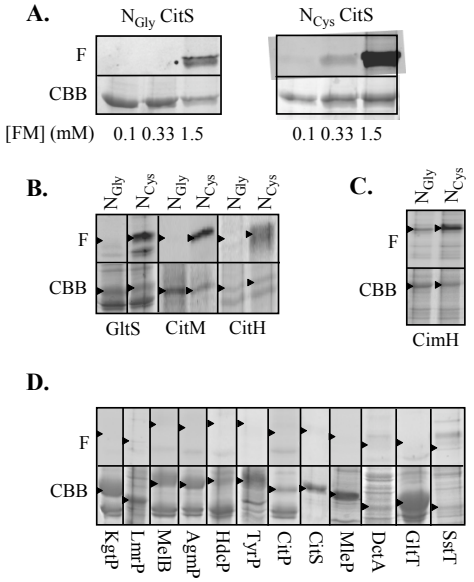


Figure 5. N-terminal localization of full-length proteins. *E. coli* SF100 cells expressing the Gly or Cys variant of the full-length proteins as indicated were treated with the indicated concentration of FM (A) or 0.1 mM FM (B,C,D) for 20 min. Following partial purification, the protein were analyzed by SDS-PAGE (see Methods). F, fluorescence image of the part of the gel containing the protein; CBB, same part of the gel after staining with Coomassie Brilliant Blue. Arrow heads indicate the positions of the partially purified proteins.

CimH of *B. subtilis* in structural class ST[3] was the only protein that showed labeling with FM in the Gly variant at the conditions used (Figure 5C). CimH contains 2 endogenous Cys residues. With the Cys variant, fluorescence intensity increased somewhat, but the ratio of fluorescence intensity over protein stain was lower than for GltS, CitM and CitH above. A firm assignment of the N-terminal location of CimH could not be made. All other transporters showed no labeling in the Gly (not shown) or Cys variant (Figure 5D) while the proteins could be identified by CBB staining. It was concluded that the N-termini of the proteins were situated in the cytoplasm. Summarizing, the transporters of structural classes ST[1], ST[2] and ST[4] have cytoplasmic N-termini. For ST[3] the results are more diverse. The 2HCT transporters (CitS, CitP and MleP) have cytoplasmic N-termini as well, while the N-termini of the ESS transporter (GltS) and the CitMHS transporters (CitM and CitH) were found in the periplasm. No firm conclusion could be drawn for CimH, a 2HCT transporter in ST[3].

4. Discussion

4.1. Overall performance

Limited experimental data may be very useful to improve membrane topology predictions by computational methods [11,13] and to discriminate between different models derived from different methods. Here, a method is presented by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method involves rapid cloning followed by conventional localization assays. The procedure was set up to allow for rapid screening of many proteins in a funnel type of approach by which proteins that do not result in positive localizations at all three positions are dropped from the screen. In this way, when a number of proteins from one family are included, an improved topology model for the family will still be obtained, even if individual proteins drop out of the scheme. A total of 16 secondary transporter proteins taken from 4 different structural classes were used to test the method. A positive result was obtained for 15 of the N-termini, all of the central positions, and for 14 of the C-termini. The localization of the N-terminus of CimH could not be obtained mainly because of background labeling of endogenous Cys residues. The localization of the C-terminus of GltS was not obtained, due to low expression of the fusion

proteins and of MleP because the results were ambiguous. It follows that a full complement of three localizations was obtained for 13 out 16 proteins, giving a drop out rate of approximately 20%. Of course, incomplete sets of localizations may still be used to check for consistency with different models.

4.2. The central loop

The use of reporter fusions to determine the membrane topology of integral membrane proteins has been widely used and, also, widely criticized. The criticism mainly focuses on the use of truncated N-terminal fragments to determine the localization of internal positions in the sequence, as used here for the central loop and does not apply when the localization of the C-terminus is determined [12,13,14]). The assumption is that the truncate itself is stable and has the same topology as in the full-length protein, i.e. the topology is not dependent on missing C-terminal parts. While in many cases the assumption appears to be valid, in specific cases erroneous assignments have been reported [30]. The internal position in the present method was selected in the middle of the protein, thereby acknowledging that many transporter proteins consist of two (homologous) domains containing the same number of TMSs, and of approximately equal length. It is assumed that the N-terminal domain in the fusion protein is a stable entity by itself and therefore less prone to erroneous results. With no further experimental information the truncation site can only be chosen based upon the number of predicted TMSs. In case of MemGen, the site may be selected based on the domain structure found in the template structure. The difference between these two cases is nicely demonstrated for the three transporters AguD, HdcP and TyrP from structural class ST[2] (Figure 4). The corresponding parts of the transporters contain 12 TMSs (HdcP has one additional TMS at the C-terminus). Therefore, the proteins would be truncated in the loop between TMS VI and VII. However, the model by MemGen was based on the high resolution structures of the LeuT and vSGLT proteins (see introduction; [10]). (In the mean time, the structural similarity between the families has been confirmed by the structure of the AdiC transporter, a homologue of HdcP [6,7]). These transporters consist out of two times 5 TMSs plus two additional segments at the C-terminus (see Figure 2) and the site of truncation was chosen between TMSs V and VI. The fusion proteins showed a more consistent result in the latter case, giving a high phoA activity in combination with a low GFP signal. Truncation of the HdcP protein after TMS VI, which includes the first TMS of the second domain, resulted in both

a high phoA activity and GFP fluorescence, suggesting a scrambled C-terminus. Apparently, truncation between the two domains results in a more stable situation. In general, the fusions of the truncated proteins in this test behaved quite well, even better than the fusions of the full-length proteins. All 16 half-proteins resulted in well-defined localizations (Figure 3), 13 in the cytoplasm and 3 at the external face of the membrane. The two full-length fusion proteins for which no localization was obtained both had their C-terminus predicted in the exterior.

4.3. The N-terminus

Determining the localization of the N-termini of the proteins is the most elaborate and cumbersome part of the screen. Following treatment of the cells with the fluorescent probe FM, identification of labeling of the engineered Cys residue at the N-terminus requires partial purification of the membrane protein. For this purpose, a His_6-tag was engineered at the N-terminus, which allows for a rapid small-scale partial purification of the proteins by affinity chromatography. Adverse affects of the presence of the His-tag at the N-terminus on the expression of the proteins may increase the dropout rate of proteins from the screen. An additional disadvantage of the labeling approach is that the absence of labeling is a negative result and it cannot be excluded that the introduced Cys residue is not accessible at the exterior of the cell rather than having a cytoplasmic location. To minimize this possibility, the Cys residue was introduced in front of the His_6-tag at the N-terminus. Then, the six histidine residues form a hydrophilic linker between the Cys residue and the hydrophobic membrane protein promoting its localization in the water phase. A further problem may arise from endogenous Cys residues that are accessible from the exterior of the cell as observed for the CimH protein in this test (Figure 5C). In general, bacteria tent to not expose Cys residues to the external surface of the cell to prevent damage to the proteins. The problem will be minimized by analyzing the amino acid sequence (of the proteins in a family) for the presence and position of endogenous Cys residues, before a selection of proteins to be included in the screen is made.

All 16 proteins in this test of 16 transporters were identified in the N_{cys} and N_{gly} variants by SDS-PAGE indicating no expression problems caused by the N-terminal His-tags. For 15 of the 16 transporters, no labeling of endogenous Cys residues was observed in the N_{gly} variant. Nevertheless, an attempt to identify labeling of the N_{cys} variant by SDS-PAGE of isolated

cytoplasmic membranes, thereby skipping the purification step, failed because of too much background labeling.

4.4. Evaluation of predictor

The topology models produced by MemGen and TMHMM were evaluated for those proteins for which the complete set of 3 experimentally determined localizations was obtained (Table 2). The models for the proteins in structural classes ST[1] and ST[2] were similar for the two methods and agreed well with the experimental data. In contrast, for each of the proteins in class ST[3], one of the predicted localizations is different in the models produced by MemGen and TMHMM.

Table 2.

Comparison[a] of experimentally determined (Exp) and predicted localizations[b] by MemGen (MG), TMHMM (TM) and TMHMM$_{fix}$ (TM$_{fix}$)[c].

Class	Protein	N-terminus				Central loop				C-terminus			
		Exp	MG	TM	TM$_{fix}$	Exp	MG	TM	TM$_{fix}$	Exp	MG	TM	TM$_{fix}$[d]
	KgtP	i	i	i	i	i	i	i	i	i	i	i	i
ST[1]	LmrP	i	i	i	i	i	i	i	i	i	i	i	i
	MelB	i	i	i	i	i	i	i	i	i	i	i	i
	AguD	i	i	i	i	o	o	o	o	i	i	i	i
ST[2]	HdcP	i	i	i	i	o	o	o	o	o	o	o	o
	TyrP	i	i	i	i	o	o	o	o	i	i	i	i
	CitH	o	o	i	i	i	i	i	i	o	o	o	o
ST[3]	CitM	o	o	i	i	i	i	i	i	o	o	o	o
	CitP	i	i	i	i	i	i	o	o	o	o	o	o
	CitS	i	i	o	o	i	i	i	i	o	o	o	o
	DctA	i	i	i	i	i	i	i	i	i	i	i	i
ST[4]	GltT	i	i	o	o	i	i	i	i	i	i	i	i
	SSTT	i	i	i	i	i	i	i	i	i	i	i	i

[a] Grayed areas indicate predicted localizations at variance with the experimental data

[b] i, cytoplasmic (in); o, external (out);

[c] TMHMM with constraint C-terminal localization

[d] Copied from column C-terminus Exp

Comparison to the experimental data shows that in all cases the prediction by MemGen is correct. Similarly, the experimental data is consistent with the MemGen model in case of the GltT protein in class ST[4] and not with the TMHMM model. Prediction for the other two proteins in class ST[4] were the same and in agreement with the experimental data. Summarizing, MemGen models agree with the experimental data in all cases (100%), while the TMHMM models agree in 8 out of 13 cases (62%) Constraining the C-terminal localization in the algorithm ($TMHMM_{fix}$; [11]) in none of the cases changed the prediction of the other two sites and, therefore, no improvement of the erroneously predicted models by TMHMM was obtained. It appears that the success rate of TMHMM is strongly dependent on the structural class with correct predictions in classes ST[1] and ST[2] and erroneous predictions in classes ST[3] and ST[4]. The low success rate for proteins in ST[3] is in agreement with observations made before (see Introduction; [21]). The transporters in class ST[3] are believed to contain two so-called pore-loops or reentrant loops [17,19,21]. Such loops are also prominent features of the transporters of class ST[4] [32] but are absent in proteins from ST[1] and ST[2]. TMHMM does not predict pore-loops and easily confuses them with transmembrane segments.

The TMHMM server was released a decade ago. During the last couple of years a number of new topology predictors for membrane proteins have been presented among which MEMSAT3, a neural network predictor that makes use of evolutionary information [39], OCTOPUS, a Hidden Markov model predictor that includes reentrant-, membrane dip- and transmembrane hairpin regions [40] and TOPCONS, a consensus predictor [41]. These new methods score considerably better with our dataset giving 100%, 92% and 85% consistency between models and experimental data for MEMSAT3, OCTOPUS and TOPCONS, respectively (Table 3).

Clearly, there has been considerable progress in the development of membrane topology prediction methods. It should be mentioned though, that models consistent with the experimental data are not necessarily the same models. In many cases, the models predicted by these predictors are still different from the MemGen models. The topology screen that is proposed here positively identifies models that are wrong. There may still be several models that are consistent with the 3 experimentally determined localizations, in which case more experiments are required to positively identify the correct model.

Table 3.

Performance of different predictors on the data set.

Predictor	ST[1]	ST[2]	ST[3]	ST[4]	All (%)
MemGen	3/3[a]	3/3	4/4	3/3	100
TMHMM	3/3	3/3	0/4	2/3	62
MEMSAT3	3/3	3/3	4/4	3/3	100
OCTOPUS	3/3	3/3	4/4	2/3	92
TOPCONS	3/3	3/3	2/4	3/3	85

[a] number of consistent topology models/number of transporters. Only those transporters were included for which a location was obtained for all three sites in the proteins.

The authors appreciate the help of Eric Geertsema of the Membrane Enzymology Group of the University of Groningen for his help with setting up the LIC system.

References

1. M. H. Saier, A functional-phylogenetic classification system for transmembrane solute transporters, *Microbiol. Mol. Rev.* **64** (2000) 354-411.
2. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, and E. Gouaux, Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters, *Nature* **437** (2005) 215-223.
3. S. Faham, A. Watanabe, G. Mercado Bessemer, D. Cascio, A. Szpecht, B. A. Hirayama, E. M. Wright, J. Abramson, The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport, *Science* **321** (2008) 810-814.
4. S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E. P. Carpenter, N. G. Rutherford, J. M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S. G. Patching, R. J. Hope, H. T. Norbertczak, P. C. Roach, S. Iwata, P. J. Henderson, A. D. Cameron, Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter, *Science* **322** (2008) 709-713.
5. S. Ressl, A. C. Terwisscha van Scheltinga, C. Vonrhein, V. Ott, C Ziegler, Molecular basis of transport and regulation in the Na+/betaine symporter BetP, *Nature* **458** (2009) 47-52.
6. Y. Fang, H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, C. Miller, Structure of a prokaryotic virtual proton pump at 3.2 A resolution, *Nature* **460** (2009) 1040-1043.
7. Xiang Gao, Feiran Lu, Lijun Zhou, Shangyu Dang, Linfeng Sun, Xiaochun Li, Jiawei Wang, and Yigong Shi, Structure and Mechanism of an Amino Acid Antiporter, *Science* **324** (2009) 1565-68.
8. J. S. Lolkema, and D. J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, *Mol. Membr. Biol.* **15** (1998) 33-42.
9. J. S. Lolkema, and D. J. Slotboom, Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins, *FEMS Microbiol. Rev.* **22** (1998) 305-322.
10. J.S. Lolkema, and D. J. Slotboom, The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-Leucine transporters, *Mol. Membr. Biol.* **25** (2008) 567-570.

11. K. Melén, A. Krogh, G. von Heijne, Reliability measures for membrane protein topology prediction algorithms, *J. Mol. Biol.* **327** (2003) 735-44.
12. D. O. Daley, M. Rapp, E. Granseth, K. Melén, D. Drew, G. von Heijne, Global topology analysis of the *Escherichia coli* inner membrane proteome, *Science* **308** (2005) 1321-3.
13. M. Rapp, D. Drew, D.O. Daley, J. Nilsson, T. Carvalho, K. Melén, J. W. De Gier and G. Von Heijne, Experimentally based topology models for *E. coli* inner membrane proteins, *Protein science* **13** (2004) 937-45.
14. H. Kim, K. Melén, M. Osterberg, G. von Heijne, A global topology map of the *Saccharomyces cerevisiae* membrane proteome, *Proc. Natl. Acad. Sci. USA* **103** (2006) 11142-7.
15. E. Granseth, D. O. Daley, M. Rapp, K. Melén, G. von Heijne, Experimentally constrained topology models for 51,208 bacterial inner membrane proteins, *J. Mol. Biol.* **352** (2005) 489-94.
16. J. S. Lolkema, and D. J. Slotboom, Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis, *J. Mol. Biol.* **327** (2003) 901-909.
17. A. J. Dobrowolski, I. Sobczak, and J. S. Lolkema, Experimental Validation of Membrane Topology Prediction by Hydropathy Profile Alignment: Membrane Topology of the Na^+-Glutamate Transporter of *Escherichia coli*, *Biochemistry* **46** (2007) 2326-2332.
18. I. Sobczak and J. S. Lolkema, The 2-hydroxycarboxylate transporter (2HCT) family: Physiology, Structure and Mechanism, *Microbiol. Mol. Biol. Rev.* **69** (2005) 665-695.
19. A. Dobrowolski, J. S. Lolkema, Functional Importance of GGXG Sequence Motifs in Putative Reentrant Loops of 2HCT and ESS Transport Proteins, *Biochemistry* **48** (2009) 7448-56
20. A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* **305** (2001) 567-580.
21. J. S. Lolkema, Domain structure and pore-loops in the 2-hydroxycarboxylate transporter family, *J. Mol. Microbiol. Biotechnol.* **11** (2006) 318-325.
22. F. Baneyx, G. Georgiou, In vivo degradation of secreted fusion proteins by the Escherichia coli outer membrane protease OmpT, *J. Bacteriol.* **172** (1990) 491-4.

23. C. Gutierrez, J.C. Devedjian, A plasmid facilitating *in vitro* construction of phoA gene fusions in *Escherichia coli*, *Nucleic Acids Res.* **17** (1989) 3999.
24. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter, *J. Bacteriol*. **177** (1995) 4121-4130.
25. E. R. Geertsma and B. Poolman, High-throughput cloning and expression in recalcitrant bacteria, *Nature Methods* **4** (2007) 705-707.
26. C. Manoil, Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions, *Methods in cell biology* **34** (1991) 61-75.
27. D. Drew, D. Sjöstrand, J. Nilsson, T. Urbig, C. N. Chin, J. W. de Gier and G. von Heijne, Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis, *Proc. Natl. Acad. Sci. USA* **99** (2002) 2690–2695.
28. B. J. Feilmeier, G. Iseminger, D. Schroeder, H. Webber and G. J. Phillips, Green Fluorescent Protein Functions as a Reporter for Protein Localization in *Escherichia coli*, *Journal of Bacteriol.* **182**(14) 4068-4076, 2000.
29. C. Manoil, and J. Beckwith, A genetic approach to analyzing membrane protein topology, *Science* **233** (1986) 1403-1408.
30. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*, *Science* **301** (2003) 610-615.
31. Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*, *Science* **301** (2003) 616-620.
32. D. Yernool, O. Boudker, Y. Jin, and E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, *Nature* **431** (2004) 811-818.
33. M. van Geest, and J. S. Lolkema, Membrane topology and insertion of membrane proteins: search for topogenic signals, *Microbiol. Mol. Biol. Rev.* **64** (2000) 13-33.
34. M. van Geest, and J. S. Lolkema, Membrane topology of the sodium ion-dependent citrate carrier of *Klebsiella pneumoniae*. Evidence for a new structural class of secondary transporters, *J. Biol. Chem.* **271** (1996) 25582-25589.
35. M. van Geest and J. S. Lolkema, Membrane topology of the Na(+)/citrate transporter CitS of *Klebsiella pneumoniae* by insertion mutagenesis, *Biochim. Biophys. Acta* **1466** (2000) 328-338.

36. I. Sobczak and J. S. Lolkema, Accessibility of cysteine residues in a cytoplasmic loop of CitS of *Klebsiella pneumoniae* is controlled by the catalytic state of the transporter, *Biochemistry* **42** (2003) 9789-9796.
37. A. Boorsma, M. E. van der Rest, J. S. Lolkema and W. N. Konings, Secondary transporters for citrate and the Mg(2+)-citrate complex in Bacillus subtilis are homologous proteins, *J. Bacteriol.* **178** (1996) 6216-6222.
38. V. S. Blancato, C. Magni, J. S. Lolkema, Functional characterization and Me ion specificity of a Ca-citrate transporter from *Enterococcus faecalis*, *FEBS J*. **273** (2006) 5121-30.
39. D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics* **23** (2007) 538-544.
40. H. Viklund and A. Elofsson, Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar, *Bioinformatics* **24** (2008) 1662-1668.
41. A. Bernsel, H. Viklund, A. Hennerdal and A. Elofsson, TOPCONS: consensus prediction of membrane protein topology, *Nucleic Acids Research* **37** (2009) Web Server issue W465-68.
42. W. Seol and A. J. Shatkin, Membrane topology model of *Escherichia coli* alpha-ketoglutarate permease by phoA fusion analysis, *J. Bacteriol*. **175** (1993), 565-567.
43. G.J. Poelarends, P. Mazurkiewicz and W.N. Konings, Multidrug transporters and antibiotic resistance in *Lactococcus lactis*, Biochim. Biophys. Acta **1555** (2002), 1-7.
44. T. Pourcher, E. Bibi, H. R. Kaback and G. Leblanc, Membrane topology of the melibiose permease of *Escherichia coli* studied by melB-phoA fusion analysis, *Biochemistry* **35** (1996), 4161-4168.
45. P. M. Lucas, V. S. Blancato, O. Claisse, C. Magni, J. S. Lolkema and A. Lonvaud-Funel, Agmatine deiminase pathway genes in *Lactobacillus brevis* are linked to the tyrosine decarboxylation operon in a putative acid resistance locus, *Microbiology* **153** (2007), 2221-2230.
46. P. M. Lucas, W. A. M. Wolken, O. Claisse, J. S. Lolkema and A. Lonvaud-Funel, Histamine-Producing Pathway Encoded on an Unstable Plasmid in *Lactobacillus hilgardii* 0006, *Applied and Environmental Microbiology* **71** (2005), 1417-1424.

47. W. A. M. Wolken, P. M. Lucas, A. Lonvaud-Funel and J. S. Lolkema, The Mechanism of the Tyrosine Transporter TyrP Supports a Proton Motive Tyrosine Decarboxylation Pathway in *Lactobacillus brevis*, *J. Bacteriol.* **188** (2006), 2198–2206.
48. B. P. Krom, R. Aardema and J. S. Lolkema, Bacillus subtilis YxkJ Is a Secondary Transporter of the 2-Hydroxycarboxylate Transporter Family That Transports L-Malate and Citrate, *J. Bacteriol.* **183** (2001), 5862–5869.
49. B. P. Krom, J. B. Warner, W. N. Konings and J. S. Lolkema, Complementary Metal Ion specificity of the Metal-Citrate Transporters CitM and CitH of Bacillus subtilis, *J. Bacteriol.* **182** (2000), 6374–6381.
50. I. Sobczak and J. S. Lolkema, The 2-Hydroxycarboxylate Transporter Family: Physiology, Structure, and Mechanism, *Microbiol. Mol. Biol. Rev.*, **69** (2005), 665–695.
51. A. Dobrowolski, I. Sobczak-Elbourne, J. S. Lolkema, Membrane topology prediction by hydropathy profile alignment: membrane topology of the Na(+)-glutamate transporter GltS, *Biochemistry* **46** (2007), 2326-32.
52. D.-J. Slotboom, W. N. Konings, and J. S. Lolkema, Structural features of the glutamate transporter family, *Microbiol. Molec. Biol. Rev.* **63** (1999), 293-307.

Chapter 3

Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification Confirmation and structural diversity

Ramon ter Horst and Juke S. Lolkema

Biochim. Biophys. Acta. **1818** (2011) 72-81

The MemGen structural classification of membrane proteins groups families of proteins by hydropathy profile alignment. Class ST[3] of the MemGen classification contains 32 families of transporter proteins including the IT superfamily. Transporters from 19 different families in class ST[3] were evaluated by the TopScreen experimental topology screening method to verify the structural classification by MemGen. TopScreen involves the determination of the cellular disposition of three sites in the polypeptide chain of the proteins, which allows for discrimination between different topology models. For nearly all transporters at least one of the predicted localizations is different in the models produced by MemGen and predictor TMHMM. Comparison to the experimental data showed that in all cases the prediction by MemGen was correct. It is concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other families in class ST[3]. The core structure of the model consists of two homologous domains, each containing 5 transmembrane segments, which have an opposite orientation in the membrane. A reentrant loop is present in between the 4th and 5th segment in each domain. Nearly all of the identified and experimentally confirmed structural variations involve additions of trans membrane segments at the boundaries of the core model, at the N- and C-termini or in between the two domains. Most remarkable is a domain swap in two subfamilies of the [st312]NHAC family that results in an inverted orientation of the proteins in the membrane.

1. Introduction

Secondary transporters are integral membrane proteins encoded by a single gene. They consist of a bundle of α-helices that are more or less perpendicular to the plane of the membrane. Despite this simple architecture, their phylogenetic diversity is enormous as evidenced by the approximately 100 gene (super)families in the Electrochemical potential-driven transporters section of the Transport Classification system (TC system) [1]. Most likely, the genetic diversity is a consequence of divergent evolution and many different families may represent a similar fold and translocation mechanism. In recent years high resolution X-ray structures have been presented of transporters from different gene families that showed the same core structure while no significant sequence similarity could be identified between the transporters. The families include the Neurotransmitter:Sodium Symporter (2.A.22 NSS) family (structure of LeuT) [2], the Solute:Sodium Symporter (2.A.21 SSS) family (vSGLT) [3], the Nucleobase:Cation Symporter-1 (2.A.39 NCS1) family (Mhp1) [4], the Betaine/Carnitine/Choline Transporter (2.A.15 BCCT) family (BetP) [5] and the Amino Acid-Polyamine-Organocation (2.A.3 APC) family (AdiC, ApcT) [6,7,8]. Structural similarity between transporters of the NSS, SSS, NCS1 and APC families was predicted by the MemGen classification system that identifies distant evolutionary relationships by hydropathy profile alignment [9,10,11]. MemGen groups families of transporters with the same global fold in structural classes. So far, 4 structural classes were defined termed ST[1], ST[2], ST[3] and ST[4]. The NSS, SSS, NCS1 and APC families are found in class ST[2]. In addition to ST[2], high resolution structures are available for class ST[1] (LacY, GlpT) [12,13] and class ST[4] (Glt_{Ph}) [14]. Class ST[3] in the MemGen classification groups 32 protein families and includes the IT superfamily [15] in the TC system [1,16]. Functional characterized members of the 32 families are secondary transporters of inorganic and organic anions and Na^+/H^+ antiporters. Most families contain exclusively transporters from prokaryotic origin, but f.i. the Divalent Anion:Na^+ Symporter (2.A.47 DASS) family (also known as SLC13) contains eukaryotic transporters for organic di- and tricarboxylate Krebs cycle intermediates as well as dicarboxylate amino acids, and inorganic sulfate and phosphate ions. No high resolution structure is available of any of the transporters in class ST[3]. Detailed studies of the Na^+-citrate transporter CitS of *Klebsiella pneumonia* and the Na^+-glutamate transporter GltS of *Escherichia coli*, members of the [st326]2HCT and [st324]ESS families in class ST[3], respectively, have

resulted in a model for the common core structure of the transporters. The core consists of two homologous domains with opposite orientation in the membrane and containing 5 TMSs and a re-entrant (or pore) loop each [17]. Members of the two families share no significant sequence similarity, but the data strongly suggested that CitS and GltS share the same fold which confirmed their assignment to the same structural class [18].

To verify the structural classification by MemGen, we have developed the TopScreen membrane topology screening method that experimentally discriminates between different topology models [19]. TopScreen involves determination of the cellular disposition of three positions of the polypeptide chain of a membrane protein using a combination of conventional techniques and was used before to confirm the distribution of a total of 16 secondary transporters over the 4 MemGen structural classes ST[1], ST[2], ST[3] and ST[4]. Here, the TopScreen approach was used to confirm the assignment of 19 transporter families to class ST[3] thereby strongly indicating that the [st326]2HCT/[st324]ESS structural core model applies to all the families in ST[3] including the families of the IT superfamily. Structural variations adopted by different (sub)families in class [ST3] are discussed.

2. Materials and Methods

2.1 Materials

Phusion DNA polymerase was obtained from Finnzymes (Espoo, Finland). T4 ligase was obtained from New England Biolabs (Frankfurt am Main, Germany). All other enzymes were obtained from Fermentas (Burlington, Canada). Mutagenic oligonucleotides were obtained from Biolegio (Nijmegen, The Netherlands), or from Operon (Ebersberg, Germany) for ligation independent cloning. *p*-Nitrophenyl phosphate (pNPP) was obtained from Sigma (Zwijndrecht, The Netherlands), fluorescein maleimide (FM) from Invitrogen (Carlsbad, United States)

2.2. Bacterial strains and growth conditions

Escherichia coli strain SF100 (*recA Δlac ΔompT*) [20] harboring the indicated pLIC vector (see below), was routinely grown in Luria Broth medium at 37 ^{o}C, with ampicillin added at a final

concentration of 50 μg/ml. Overnight cultures were diluted 30-fold in 3 mL of fresh medium and when the optical density measured at 660 nm (OD_{660}) reached a value between 0.6 and 0.8, arabinose was added at a final concentration of 0.002-0.05 % (w/v) to induce protein production from the plasmids. Following growth for another 1.5-2 h, cells were harvested by centrifugation in a table top centrifuge operated at 4 ^{o}C. Cells were resuspended in the indicated buffer and kept on ice until use.

2.3. Ligation Independent Cloning.

All genetic manipulations were as described before [19]. Briefly, a synthetic double stranded piece of DNA (the LIC cassette) was inserted downstream of the arabinose promoter in the commercial pBAD24 vector (Invitrogen) [21]. The nucleotide sequence of the sense strand was 5'-C ATG GGT *CAT CAT CAC CAT CAC CA**T*** **TTA AA**T AGT GGT GTG GTA C- 3' in which the initiation codon was underlined, the sequence coding for a His_6-tag set in italics and a *SwaI* restriction site in bold. The resulting vector with a gene encoding *phoA* or *gfp* inserted in frame with the initiation codon was used to construct alkaline phosphatase or green fluorescence protein fusion proteins, and was termed pLIC1 or pLIC2 respectively. Replacing the *phoA* fragment by a double stop codon resulted in vector pLIC3 which is used to produce His-tagged proteins. The second codon of the cassette in pLIC3, GGT (Gly) was mutated into TGT (Cys) by site directed mutagenesis rendering vector pLIC4 which is used for N-terminal localizations.
Ligation Independent Cloning (LIC) was done essentially as described [22]. The LIC vectors pLIC1-4 were linearized by *Swa*I digestion. Transporter genes were amplified using forward and backward primers containing 5' flanking regions corresponding to the nucleotide sequences upstream and downstream of the *Swa*I site in the LIC cassette as follows: 5'-CATGGGTCATCATCACCATCACCATTTG.....-3' and 5'-TACCACACCACTATTTTG.....-3', respectively. For each individual gene one forward and two reverse primers were used to amplify full length and half length transporter genes. Single-stranded overhangs of the PCR products and vectors were generated using T4 DNA polymerase in the presence of dGTP (vector) and dCTP (PCR product). The complementary overhangs of PCR product and vector annealed upon mixing, after which the resulting heteroduplexes were transformed to *E. coli* SF100.

2.4. GFP and PhoA assays

2.4.1. PhoA assay.

*C*ells from 0.5 ml of culture of *E. coli* SF100 cells harboring a pLIC1 vector carrying the indicated insert were washed once, and subsequently resuspended in 1 M Tris-HCl pH 8.0 buffer after which the OD_{660} was measured. Following incubation for 5 min at 37 ^{o}C, 200 µL of a 1.5 mM *p*-nitrophenyl phosphate (pNPP) solution was added and the suspension was incubated until a yellow color developed. Then, the cells were spun down and the absorption was read at 420 and 550 nm using a Hitachi U-1100 spectrophotometer. The phosphatase activity was calculated in Miller units [23].

2.4.2. GFP assay.

Cells of *E. coli* SF100 harboring a pLIC2 vector carrying the indicated insert were washed once and, subsequently, resuspended in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 15 mM EDTA to a final OD_{660} of 0.2. Following incubation for 30 min at room temperature, a 150 µL cell suspension was transferred into a precision cell (Hellma, Quartz SUPRSIL) and GFP fluorescence emission intensity at 508 nm was measured at an excitation wavelength of 470 nm using an AMINCO Bowman Series 2 Luminescence Spectrometer. Background signals caused by light scattering were estimated from cells harboring the pLIC3 vector without insert [24].

2.4.3. Evaluation of data.

Mean values and standard deviations were calculated from at least three independent measurements. PhoA and GFP activities of the cells were normalized by the mean PhoA activity of all positive PhoA fusions (>100 Miller units) and all positive GFP fusions (> 0.5 emission units), respectively. The logarithm of the ratio of the normalized PhoA and GFP activities was calculated for each full-length and half protein to obtain a measure for the cellular localization of the fusion point. A positive value corresponds to a periplasmic localization, a negative bar to a cytoplasmic localization. The length of the bar indicates the significance of the localization and would be independent on the expression levels [19]. Values >2.5 or <-2.5 were arbitrarily set to 2.5 and -2.5, respectively.

2.5. Labeling studies

E. coli SF100 cells harboring pLIC4 or pLIC3 vectors carrying the indicated insert were washed once and, subsequently, resuspended in ice-cold 50 mM potassium phosphate buffer pH 7.0. Cells from 200 ml of culture were resuspended in 1 ml of buffer. A 5 mM solution of fluorescein 5-maleimide (FM) was freshly prepared by diluting a 50 mM solution in DMSO with 50 mM potassium phosphate buffer pH 7.0. FM was added to the cell suspension to a final concentration of 100 μM. After incubation for 20 min at room temperature, excess FM was quenched by adding 2 mM dithiothreitol (DTT). Cells were washed three times with 50 mM potassium phosphate buffer pH 7.0, and disrupted by sonication. Following removal of debris, membranes were collected from the supernatant by centrifugation for 25 min at 80,000 rpm in a Beckman TLA100.2 rotor at 4 ^{o}C. The pellet was solubilized in 1 mL 50 mM potassium phosphate buffer pH 8.0 containing 400 mM NaCl, 20% glycerol, 10 mM imidazole, 1mM DTT and 2% Triton X-100 and left on ice for 30 min after which undissolved material was removed by centrifugation at 90,000 rpm for 25 min at 4 ^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (25 μL bed volume) equilibrated in 50 mM potassium phosphate buffer pH 8.0 containing 600 mM NaCl, 20% glycerol, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100, and incubated overnight at 4^{o}C under continuous shaking. The resin was pelleted by brief centrifugation in a table top centrifuge and the supernatant was removed. The resin was washed with 0.5 mL of equilibration buffer containing 300 mM NaCl and 20-40 mM imidazole. The protein was eluted with 25 μL of the same buffer at pH 7.0 and containing 200 mM imidazole. Subsequently, samples were loaded onto a 12% sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE). Following electrophoresis, in-gel fluorescence was recorded using a Fujifilm LAS-4000 luminescent image analyzer, and the gel was stained with Coomassie Brilliant Blue (CBB).

3. Results

3.1. TopScreen and selection of ST[3] transporters

The TopScreen approach involves the determination of the cellular disposition of three positions of a membrane protein, the N- and C-termini and a position in the central loop that is measured

by the localization of the C-terminus of the half-protein [19]. The locations of the C-termini of full-length proteins and half-proteins are determined using complementary reporter fusion techniques. Alkaline phosphatase fused at the C-terminus results in an active enzyme when located in the periplasm, while fusion of GFP results in high fluorescence only when located in the cytoplasm. The N-termini are determined by the accessibility of an introduced cysteine residue by the fluorescent sulfhydryl reagent fluorescein maleimide. Table 1 lists a total of 38 proteins from structural class ST[3] of the MemGen classification that have been evaluated by the TopScreen method. Most of the proteins have been functionally characterized as secondary transporters. All of them are from bacterial origin, originating from the phyla proteobacteria and firmicutes. The proteins are distributed over 19 different families. In general no significant sequence similarities can be detected between transporters from different families. Transporters in the [st301]CITMHS, [st324]ESS and [st326]2HCT families were evaluated before [19] and were included for completeness throughout this study. Detailed topology models are available for the [st324]GLTS and [st326]2HCT families. The IT superfamily [15] from the Transport Classification system [1] (TC system) is represented by 14 different families. The [st310]Ato, [st314]AbgT, [st315]AAE, [st324]ESS, and [st326]2HCT families are not found in the IT superfamily.

Table 1. ST[3] transporter proteins evaluated by TopScreen

MemGen subfamily	Transport protein	TC superfam/ family	Organism	Lineage[a]	Function	Half-protein[b]	Reference
[st301]CitMHS1	CitHbsub	IT/CitMHS	*Bacillus subtilis*	B-F-b	Ca^{2+}-Citrate:H^+-symporter	R202	25
	CitMsbub				Mg^{2+}-Citrate:H^+-symporter	E208	25
[st302]ArsB1	ARSBecol	IT/ArsB	*Escherichia coli*	B-P-c	Arsenical pump sub-unit	F224	26
[st302]ArsB5	YBIRecol				-	-	-
[st303]DASS1	YBHIecol	IT/DASS	*Escherichia coli*	B-P-c	-	K250	-
	TtdTecol				Predicted tartrate/succinate antiporter	Q277	27
[st304]GntP1	GntPecol	IT/GntP	*Escherichia coli*	B-P-c	Gluconate transporter	F210	28
	IDNTecol				L-idonate and D-gluconate transporter	-	
[st305]Dcu1	DcuBecol	IT/Dcu	*Escherichia coli*	B-P-c	Anaerobic C4-dicarboxylate transporter	F206	29, 30
	DcuAecol				C4-dicarboxylate transporter	E206	31
[st306]NhaB1	NHABecol	IT/NhaB	*Escherichia coli*	B-P-c	Na^+/H^+ antiporter	D284	32
	NHABpflu		*Pseudomonas fluorescens*		Na^+/H^+ antiporter	D290	-
[st307]TRAP-T1a	YGIKecol	IT/TRAP-T	*Escherichia coli*	B-P-c	-	H219	-
	YIANecol		*Escherichia coli*	B-P-c	-	Q201	-
	YP349185pflu		*Pseudomonas fluorescens*	B-P-c	-	-	-
	YP428757rrub		*Rhodospirillum rubrum*	B-P-a	-	K203	-
[st309]DCUC1	DCUCecol	IT/DcuC	*Escherichia coli*	B-P-c	Anaerobic C4-dicarboxylate transporter	R220	33
	DCUDecol				-	I226	-
[st310]Ato1	ATOEecol	Ato	*Escherichia coli*	B-P-c	Short chain fatty acid transporter	P209	-
	YP350496pflu		*Pseudomonas fluorescens*		-	K228	-
[st311]AIT-A1	ZP03593717bsub	IT	*Bacillus subtilis*	B-F-b	-	A205	-
[st312]NhaC1	NHACbsub	IT/NhaC	*Bacillus subtilis*	B-F-b	Na^+/H^+ antiporter	Q224	34
[st312]NhaC1	MLENbsub		*Bacillus subtilis*	B-F-b	H^+-Malate/Na^+-Lactate antiporter	E224	35
[st312]NhaC2	NP149303cace		*Clostridium acetobutylicum*	B-F-c	-	E222	-
[st312]NhaC2	NP207738bsub		*Bacillus subtilis*	B-P-b	-	Y230	-
[st312]NhaC4	YUIFbsub		*Bacillus subtilis*	B-F-b	-	K216	-
[st313]AIT-B1	YCGAbsub	IT	*Bacillus subtilis*	B-F-b	-	D176[c]	-
[st314]AbgT1	AbgTecol	AbgT	*Escherichia coli*	B-P-c	aminobenzoyl-glutamate transporter	R239	36
[st315]AAE2	AspTstrt	AAE	*Streptococcus thermophilus*	B-F-l	Putative aspartate/alanine exchanger	E223	37
[st320]AIT-F1	YHFAbsub	IT	*Bacillus subtilis*	B-F-b	-	E223	-
[st322]AIT-H1	YJCLbsub	IT	*Bacillus subtilis*	B-F-b	-	E192	-
[st324]ESS1	GltSecol	ESS	*Escherichia coli*	B-P-c	Na^+:Glutamate symporter	Q197	18
[st325]LctP1	GLCAecol	IT/LctP	*Escherichia coli*	B-P-c	Glycolate transporter	E222/E275[d]	38
	YVFHbsub		*Bacillus subtilis*	B-F-b	-	E222/E280[d]	-
	LldPecol		*Escherichia coli*	B-P-c	L-Lactate transporter	-/R276	38
[st326]2HCT1	CimHbsub	2HCT	*Bacillus subtilis*	B-F-b	L-Malate:Citrate-H^+ symporter	E261	39
	CitPlmes		*Leuconostoc mesenteroides*	B-F-l	Citrate/Lactate exchanger	V269	17
	CitSkpne		*Klebsiella pneumonia*	B-P-c	Citrate:Na^+ symporter	S251	17

[a] B-P-a/c/e: Bacteria, Proteobacteria, alpha/gamma/epsilon subdivision; B-F-l: Bacteria, Firmicutes, lactobacillales; B-F-b: Bacteria, Firmicutes, bacillales; B-F-c: Bacteria, Firmicutes, clostridia

[b] Last residue in N-terminal half-protein [c] Last residue in N-terminal half-protein of N-terminally truncated version of YCGAbsub [d] Two half proteins were analyzed (see Fig. 1B)

3.2. C-terminal location of full-length and half proteins

Genes encoding the full-length and half-proteins of each of the transporters in Table 1 were cloned into the appropriate vectors resulting in PhoA and GFP fusion proteins. The C-terminal residue of each half-protein was indicated in Table 1. Alkaline phosphatase activity of cells expressing PhoA fusions was plotted against the GFP fluorescence of cells expressing the corresponding GFP fusion (Figure 1A and B, left). The data points formed two well separated groups falling along the two axes. High alkaline phosphatase activity and low GFP fluorescence corresponds to a periplasmic localization of the fusion point whereas low alkaline phosphatase activity and high GFP fluorescence corresponds to a cytoplasmic localization. It follows that, with the exception of four, all full-length proteins would have the C-terminus in the periplasm (Figure 1A, left), and with the exception of two, all half proteins would have their C-termini in the cytoplasm (Figure 1B, left). Therefore, the C-terminal halves of most of the proteins contain an odd number of TMSs.

The normalized ratios of the PhoA and GFP signals of the individual proteins (Figure 1A and B, right; see also Methods) which more or less is independent of the expression levels of the fusion proteins [19] show that three of the full-length proteins with cytoplasmic C-termini are found in the family of Na^+/H^+ antiporters [st312]NHAC, suggesting structural diversity in this family (Figure 1A, right). MLENbsub [35], a H^+-malate/Na^+-lactate antiporter, and NHACbsub [34], a Na^+/H^+ antiporter, both from *Bacillus subtilis* are found in subfamily [st312]NHAC1 and both have a cytoplasmic C-terminus. In subfamily [st312]NHAC2, the putative transporter NP207738 of *Heliobacter pylori* ends in the cytoplasmic as well, but NP149303 of *Clostridium acetobutylicum* in the same subfamily ends in the periplasm.

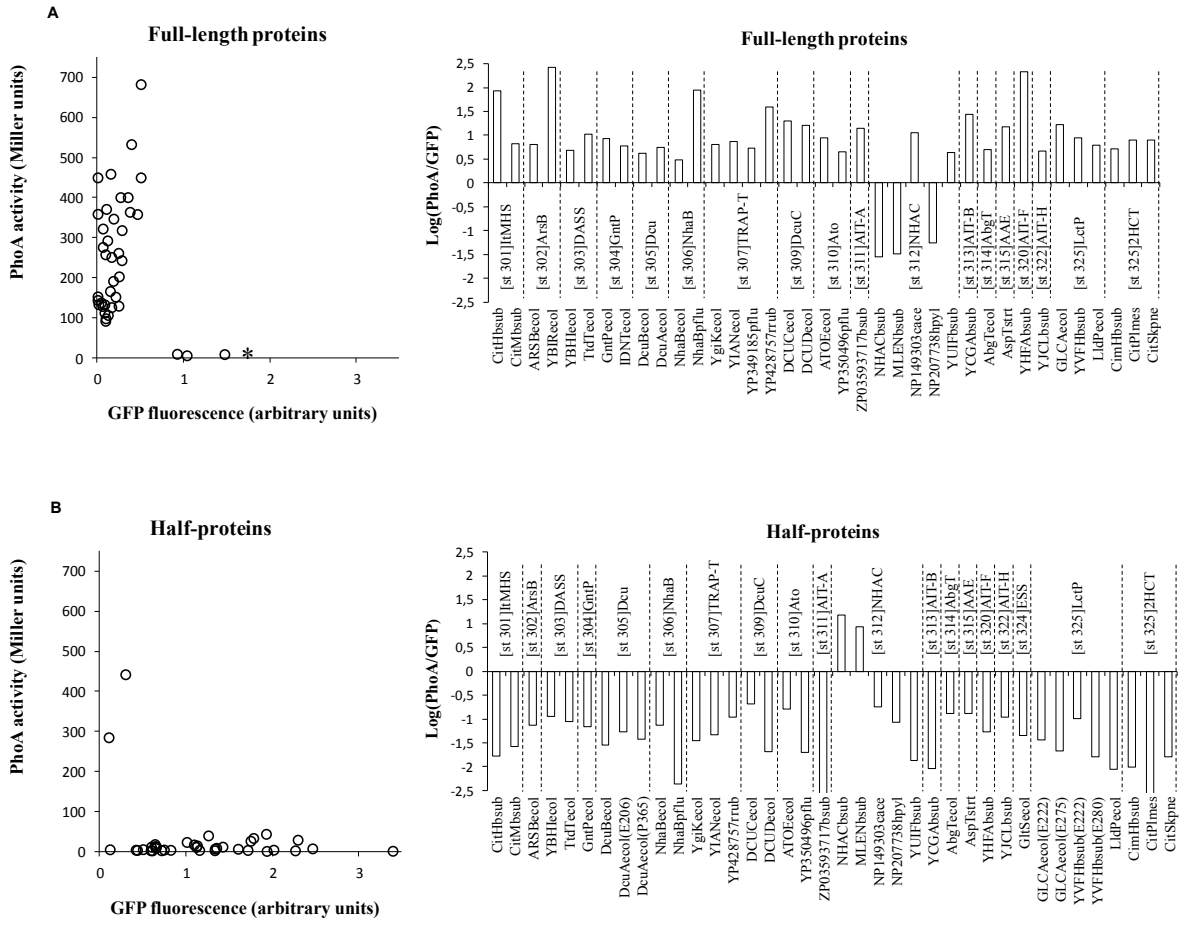


Figure 1. C-terminal localization of full-length (A) and half-proteins (B). PhoA activity *versus* GFP fluorescence intensities (left panels) and logarithm of the ratio of the normalized PhoA and GFP signals (right panels) for full length proteins and half-proteins of the ST[3] transporters listed in Table 1. *, C-terminally truncated version of DcuA of *E. coli.*

Finally, YUIFbsub of *B. subtilis* in subfamily [st312]NHAC4 also has a periplasmic C-terminus. The two exceptions with the C-termini of the half proteins in the periplasm are found in family [st312]NHAC as well (Figure 1B, right). MLENbsub and NHACbsub of *B. subtilis* in subfamily [st312]NHAC2 that were exceptional by having their C-terminus in the cytoplasm, again are exceptional by having the C-terminus of the half protein in the periplasm. These results will be discussed in further detail below.

The C_4–dicarboxylate transporter DcuA of *Escherichia coli* in family [st305]Dcu is found in the protein databases as a protein consisting of 371 and 433 residues of which the shorter is most

likely a misannotated sequence truncated at the C-terminus. The C-terminus of the shorter sequence, DcuAecol*, was located in the cytoplasm (Figure 1A) indicating an odd number of TMSs in the C-terminal 62 residues of DcuAecol.

The hydropathy profile of the transporters in the lactate transporter family [st325]LctP reveals a hydrophobic region in the central part of the protein. Half proteins of the glycolate transporter GLCAecol [38] of *E. coli* and YVFHbsub of *B. subtilis* were constructed by truncating the sequence just before and after the hydrophobic region (see Table 1). For both proteins, both versions of the half proteins ended in the cytoplasm (Figure 1B), suggesting two TMSs in the central hydrophobic regions.

3.3. N-terminal location of full-length proteins

For each of the transporters in Table 1 two constructs were made to determine the location of the N-terminus, one variant with a Cys and one with a Gly residue at position 2, immediately upstream of the His_6-tag. Treatment of cells expressing the two variants of the transporters with 0.1 mM of the fluorescent sulfhydryl reagent fluorescein maleimide (FM) for 20 minutes and at room temperature was used to discriminate between a cytoplasmic and periplasmic location of the N-terminus of the proteins. Under these conditions the thiol reagent is membrane impermeable, and only labels Cys residues exposed to the external medium [19]. Determination of the localization of the N-termini is the more elaborate and cumbersome of the three positions in the TopScreen approach. Main problems are the labeling of endogenous Cys residues and lack of or low expression of either of the variants. The goal was to obtain one N-terminal localization in each of the nineteen families, which succeeded for all but one family.

The N-terminus of the transporters from nine different families were clearly located in the periplasm, as was observed before for CitM of *B. subtilis* in family [st301]CitMHS and GltS of *E. coli* in [st324]ESS (Figure 2A) [19]. Following treatment of the cells with fluorescein maleimide and Ni^{2+}-NTA affinity purification of the His-tagged transporters from the membranes, fluorescence imaging after SDS-PAGE showed clear labeling of the Cys variant, while the Gly variant was not labeled (Figure 2A). Coomassie Brilliant Blue staining of the gel revealed more or less the same protein levels in the membranes. It follows that these transporters expose their N-termini at the exterior of the cell surface. Transporters from three additional families did not

result in labeling of either the Gly or Cys variant as was observed before for CitS of *K. pneumoniae* in the [st326]2HCT family (Figure 2B). Both variants were clearly produced in the membrane and it was concluded that the N-termini of YBHIecol in [st303]DASS, the Na^+/H^+ antiporter NhaBecol [32] in [st306]NhaB and YCGAbsub in [st313]AIT-B reside in the cytoplasm where they are inaccessible to FM. The latter protein, YCGAbsub of *B. subtilis* is found in the protein databases containing 396 and 477 residues of which the shorter sequence is, most likely, a misannotated sequence truncated at the N-terminus. The N-terminus of the truncated sequence YCGAbsub* was exposed to the periplasm (Figure 2A) suggesting that the N-terminal 81 residue region of YCGAbsub contains an odd number of TMSs .

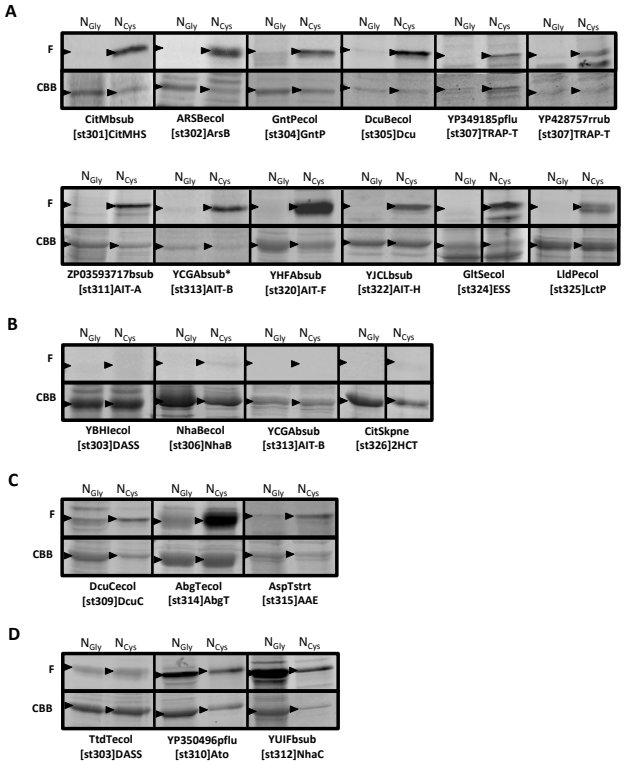


Figure 2. N-terminal localization of full-length proteins. *E. coli* SF100 cells expressing the Gly or Cys variant of the full-length proteins as indicated were treated with 0.1 mM FM for 20 min. Following partial purification, the proteins were analyzed by SDS-PAGE (see Methods). The proteins were grouped by similar results (A-D). YCGAbsub*, N-terminally truncated version of YCGAbsub of *B. subtilis*. F, fluorescence image of the part of the gel containing the protein; CBB, same part of the gel after staining with Coomassie Brilliant Blue. Arrow heads indicate the positions of the partially purified proteins.

Several of the transporters from different families showed labeling in the Gly variant, indicating the accessibility of endogenous Cys residues from the periplasmic side of the membrane (Figure 2C,D). In case of the C4-dicarboxylate transporter DCUCecol [33] in [st309]DcuC, the aminobenzoyl-glutamate transporter AbgTecol [36] in [st314]AbgT and the aspartate/alanine exchanger AspTstrt [37] of *Streptococcus thermophilus* in [st315]AAE the labeling intensity of the Gly variants was modest and the ratio of fluorescence intensity over protein stain of the bands was considerably higher for the Cys than the Gly variants indicating labeling of the N-terminal Cys residue and a periplasmic location (Figure 2C). It is concluded that the N-termini of the Cys variant of these transporters are exposed at the exterior of the cell. The Gly and Cys variants of the tartrate/succinate exchanger TtdTecol [27] in [st303]DASS and YP350496pflu in [st310]ATO showed similar ratios of fluorescence intensities over protein stain, indicating no additional labeling in the latter variant and a cytoplasmic location of the introduced Cys residue (Figure 2D). The Cys variant of YUIFbsub in [st312]NHAC showed high fluorescence intensity over protein stain, which generally indicates a periplasmic localization (see Panel A). However, a particular strong labeling was observed for the Gly variant of YUIFbsub which in combination with the low expression level of the Cys variant makes it difficult to evaluate the ratios of fluorescence intensity over protein stain rendering the localization inconclusive.

3.4. Comparison TopScreen results and MemGen predictions

Together with the results from a previous study [19], the TopScreen topology screen resulted in a complete set of three localizations for 20 proteins covering 17 families in structural class ST[3] in the MemGen classification (Table 2). In addition, combining the results from different proteins in the [st325]LctP family resulted in a complete set for this family which leaves the [st312]NHAC family as the only family for which only the location of the central loop and C-terminus could be determined. The proteins in thirteen families follow the out-in-out (o/i/o) order for the position of the N-terminus, central loop and C-terminus, respectively, which is according to the putative core structure of class ST[3] consisting of two domains with 5 TMS each with the connecting central loop in the cytoplasm. Included in this group is the [st325]LctP family, even though the hydrophobic properties of the central loop suggest additional TMSs in this region (see below). Transporters from five families have a cytoplasmic N-terminus (i/i/o) which would be in line with

the membrane topology of CitS of *K. pneumoniae* ([st326]2HCT) that has an additional TMS N-terminal to the core structure. Transporters of the [st312]NHAC family have more complicated topologies that will be discussed below.

Table 2. Comparison[a] of experimentally determined (TopScreen) and predicted localizations[b] by MemGen and TMHMM.

MemGen sub-family	Transport protein	Topscreen	MemGen	TMHMM
[st301]CitMHS1	CitHbsub	o/i/o	o/i/o	i/i/o
	CitMbsub	o/i/o	o/i/o	i/i/o
[st302]ArsB1	ARSBecol	o/i/o	o/i/o	i/i/o
[st302]ArsB5	YBIRecol	-/i/o	i/i/o	i/i/o
[st303]DASS1	YBHIecol	i/i/o	i/i/o	i/i/o
	TtdTecol	i/i/o	i/i/o	i/i/i
[st304]GntP1	GntPecol	o/i/o	o/i/o	o/o/i
	IDNTecol	-/-/o	o/i/o	i/i/i
[st305]Dcu1	DcuBecol	o/i/o	o/i/o	i/o/o
	DcuAecol	-/i/o	o/i/o	o/i/i
[st306]NhaB1	NHABecol	i/i/o	i/i/o	o/i/o
	NHABpflu	-/i/o	i/i/o	i/i/i
[st307]TRAP-T1a	YGIKecol	-/i/o	o/i/o	i/o/o
	YIANecol	-/i/o	o/i/o	o/i/i
	YP349185pflu	o/-/o	o/i/o	i/i/o
	YP428757rrub	o/i/o	o/i/o	i/i/o
[st309]DCUC1	DCUCecol	o/i/o	o/i/o	o/o/o
	DCUDecol	-/i/o	o/i/o	o/o/i
[st310]Ato1	ATOEecol	-/i/o	i/i/o	o/i/i
	YP350496pflu	i/i/o	i/i/o	o/i/o
[st311]AIT-A1	ZP03593717bsub	o/i/o	o/i/o	i/i/o
[st312]NhaC1	NHACbsub	-/o/i	i/o/i	i/i/i
[st312]NhaC1	MLENbsub	-/o/i	i/o/i	i/o/i
[st312]NhaC2	NP149303cace	-/i/o	o/i/o	i/o/i
[st312]NhaC2	NP207738hpyl	-/i/i	o/i/i	i/o/o
[st312]NhaC4	YUIFbsub	-/i/o	o/i/o	o/i/i
[st313]AIT-B1	YCGAbsub	i/o/i	i/o/i	i/i/o
[st314]AbgT1	AbgTecol	o/i/o	o/i/o	i/i/o
[st315]AAE2	AspTstrt	o/i/o	o/i/o	o/i/o
[st320]AIT-F1	YHFAbsub	o/i/o	o/i/o	i/i/o
[st322]AIT-H1	YJCLbsub	o/i/o	o/i/o	o/i/i
[st324]ESS1	GltSecol	o/i/o	o/i/o	o/i/i
[st325]LctP1	GLCAecol	-/i/i/o	o/i/i/o	i/i/o/o
	YVFHbsub	-/i/i/o	o/i/i/o	o/i/i/o
	LldPecol	o/-/i/o	o/i/i/o	o/o/i/o
[st326]2HCT1	CimHbsub	-/i/o	i/i/o	i/i/i
	CitPlmes	i/i/o	i/i/o	i/o/o
	CitSkpne	i/i/o	i/i/o	o/i/o

[a] Grayed areas indicate predicted localizations at variance with the experimental data.

[b] i, cytoplasmic (in); o, external (out).

Table 2 gives the predicted localizations of the three positions of the proteins by MemGen and by TMHMM, a widely used membrane topology predictor [40]. For nearly all transporters at least one of the predicted localizations is different in the models produced by MemGen and TMHMM.

Comparison to the experimental data shows that in all cases the prediction by MemGen is correct. While TMHMM predicts the location of the central loop mostly in the cytoplasm in line with the experimental results, the number of TMSs in the two halves of the proteins is often at variance with the experiments which, as noted before [17,18], would be due to the presence of the reentrant loop structures in the two domains.

3.5. Diversity of membrane topology in class ST[3]

A schematic representation of the predicted structural models for the different families in ST[3] is shown in Figure 3. The core structure, based on extensive studies of the membrane topology of GltS of *E. coli* ([st324]ESS family) and CitS of *K. pneumoniae* ([st326]2HCT), is represented by structure A. It consists of two domains with 5 TMSs each (5+5 topology) that are oppositely oriented in the membrane (inverted topology). In between the 4^{th} and 5^{th} segment in each domain a reentrant loop is found. In 10 families of class ST[3] the core structure is the topology adopted by most, if not all, transporters (Figure 3, structure A). Deviations from the core structure were identified when the alignment of the family hydropathy profile of a family with the profile of one of these families showed additional hydrophobic regions or, less frequently, lacked hydrophobic regions [10]. Alignment of the family profiles of the [st324]ESS family representing the core structure and the [st326]2HCT was shown before and clearly revealed an additional TMS at the N-terminus of the proteins in the latter family (1+5+5 topology; structure C) [18].

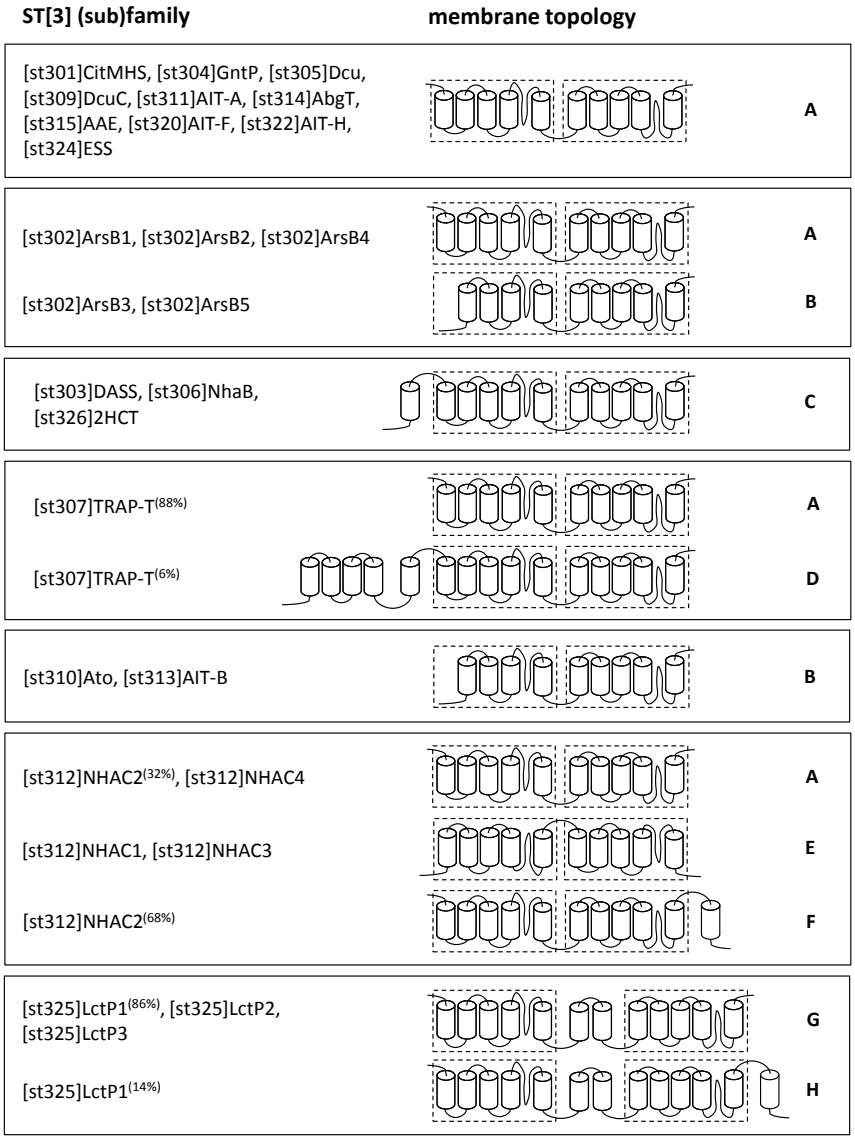


Figure 3. Membrane topology models of structural class ST[3]. See text for explanation. Boxed parts represent domains.

Figure 4A shows the additional N-terminal TMS of the [st326]2HCT family by alignment with the [st305]Dcu family. The additional TMS was also identified in the family profiles of the [st303]DASS and [st306]NhaB families and is in agreement with the TopScreen results (Table 2). The alignment in Figure 4A also shows that the locations of the C-terminus of the C-terminally truncated and full-length DCUAecol, cytoplasmic and periplasmic, respectively, are in agreement with the model even though the hydropathy profile clearly shows two hydrophobic regions in between the two positions. One of the two regions corresponds to the reentrant loop in the model.

The family profiles of the [st310]Ato and [st313]AITB families indicated that the first TMS of the core structure was missing yielding a 4+5 topology (structure B). The N-terminal localization of the N-terminally truncated YCGAbsub* protein in the periplasm would be in agreement with this model (Figure 4B). The same topology is found in the [st302]ArsB family, but only in two subfamilies. The remaining three subfamilies fold according to core structure A (Figure 3). Therefore, different transporters in one family have different numbers of TMSs.

The transporters of the [st307]TRAP-T family are exceptional as they are believed to be part of a complex that forms a transport system that appears to be a hybrid of a secondary and a ABC type of transport system [41]. The complex is believed to consists of the transporter subunit M, a smaller integral membrane protein subunit Q containing 4 TMSs and a periplasmic binding protein P. Subunits M are found in the IT superfamily and in the MemGen model they fold into the core structure of ST[3] (Figure 3, A). In a small fraction of the transporters (~6%), the small subunit Q is fused to the N-terminus of the M subunit via a linker TMS resulting in topology D (4+1+5+5 topology). Even smaller fractions of the M subunits contain insertions of a pair of TMSs in between the two domains or between the second and third TMS of the C-domain (not indicated).

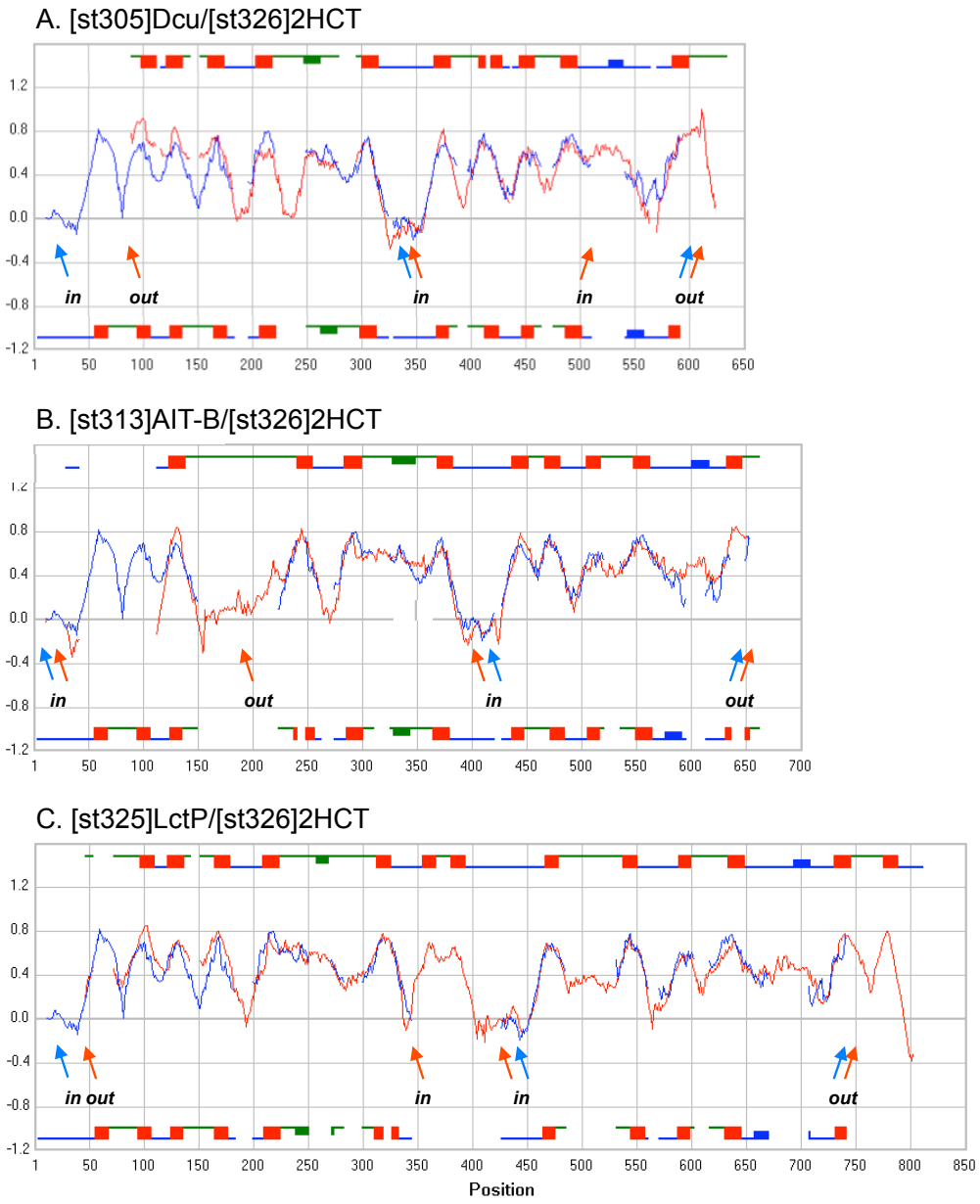


Figure 4. Family averaged hydropathy profile alignments. Alignment of the family averaged hydropathy profiles of [st326]2HCT (blue) and (A) [st305]Dcu (red), (B) [st313]AIT-B (red) and (C) [st325]LctP(red). Red and blue arrows indicate the positions and outcome of the experimental TopScreen localizations (in/out) of one or more members of the corresponding families (see also Figure 1B,C and 2A,B). Topology models are depicted above ([st305]Dcu, [st313]AIT-B and [st325]LctP), or below ([st326]2HCT) the corresponding alignments. Red squares, TMSs; green, external loops; blue, cytoplasmic loops; thickened loop regions, reentrant loops. Number of typical sequences [9] included in the family hydropathy profiles were 45, 31, 67 and 62 for [st326]2HCT), [st305]Dcu, [st313]AIT-B and [st325]LctP, respectively. The Structure Divergence Score (SDS), which is a measure of the similarity of the profiles after alignment [9], was 1.23 (A), 0.793 (B), and 0.751 (C).

Alignment of the family profile of the [st325]LctP family and the [st326]2HCT family (Figure 4C) shows that the hydrophobic region in the central loop is not part of the core structure but, in line with the experimental results, represents two additional TMSs. The pair of TMSs would be inserted in between the two domains of the core structure (5+2+5; structure G), a feature that is present in all transporters in all three subfamilies. A minority of about 15% of the transporters in subfamily [st325]LctP1 has an additional TMS at the C-terminal end of the core structure (5+2+5+1; structure H).

The most diverse membrane topologies are found for the transporters in family [st312]NHAC. Hydropathy profile alignments show that the transporters in subfamily [st312]NHAC4 and a significant fraction of the transporters in [st312]NHAC2 fold according to the core structure A in line with the experimental evidence for YUIFbsub and NP149303cace, respectively. The predominant topology (~70%) in the [st312]NhaC2 subfamily is the core structure plus an additional TMS at the C-terminus (structure F) which was experimentally confirmed by the C-terminal localization of full length NP207738hpyl of *Heliobacter pylori* (Figure 1A). The [st312]NHAC family is the only family with transporters that have the central loop in the periplasm rather than the cytoplasm (Figure 1). This is caused by a domain swap that was reported before [16]. The domain swap causes a topology inversion and leaves the N- and C-termini in the cytoplasm (structure E). The results of the TopScreen analyses of the C-terminal localizations of transporters NHACbsub and MLENbsub of *Bacillus subtilis* in subfamily [st312]NhaC1 are in agreement with the domain swap. Sequence analysis showed the same domain swap in subfamily [st312]NHAC3, while subfamilies [st312]NAHC2 and [st312]NHAC4 follow the 'normal' order of the two domains.

4.Discussion

4.1. Validation of structural classification

Structural class ST[3] of the MemGen classification groups 32 families containing 59 subfamilies that are believed to share the same fold [16]. Class ST[3] includes in addition to the IT superfamily [15] of the TC classification system [1] several other families indicating that hydropathy profile analysis detects more distant evolutionary relationship between (membrane)

proteins than sequence analysis. Before, detailed studies of CitS of *Klebsiella pneumonia*, a member of the [st326]2HCT family, and GltS of *Escherichia coli*, a member of the [st324]ESS family, provided strong support for the same fold of the transporters in these two ST[3] families that showed no sequence identity whatsoever [17,18]. TopScreen was developed to allow for discrimination between different membrane topology models of many more families by strategically selecting three positions in the polypeptide chain for experimental determination of their localization inside or outside the cell. Here, TopScreen was used to validate the structural classification of 19 secondary transporter protein families in class ST[3]. In the case of ST[3] families, the presence of putative reentrant loops in the MemGen structural model was particularly helpful as secondary structure prediction methods often mistake reentrant loops for TMSs which results in inversion of the topology of adjacent segments. For most proteins, predictor TMHMM placed the central loop in the cytoplasm, which is in agreement with the experimental data (Table 2). However, the presence of the putative reentrant loops in the two halves of the proteins make a strong case for the MemGen topology since the location of either or both of the N- and C-termini predicted by TMHMM are at variance with the TopScreen results. In conclusion, the present study suggests that the 'core' structure identified in the [st324]GltS and [st326]2HCT families is also valid for the IT superfamily transporters and other families in class ST[3].

4.2. Comparison validated and previously reported topology models

A number of topology models of ST[3] transporters have been reported previously. ArsB of *E. coli* is a member of the [st302]ArsB family and belongs to the IT superfamily. A variety of *lacZ*, *phoA* and *blaM* gene fusions resulted in a topological model consisting of 12 membrane-spanning regions, with the central loop and the N- and C- termini localized to the cytoplasm [42]. The model is at variance with the TopScreen results and the MemGen model (Figure 3, model A). Models reported for the C4-dicarboxylate transporter DcuA of *E. coli*, ([st305]Dcu) [43] and the aspartate:alanine antiporter AspT of *Tetragenococcus halophilus* ([st315]AAE) [44] reveal two times 5 TMSs with the N- and C-termini outside the cell which is in agreement with the TopScreen results. In both models the reentrant loops in the two halves were absent (Figure 3, Model A). Similarly, the cellular localization of the N- and C-terminus and central loop of the

topology model reported for the Na^+/H^+ antiporter NhaB of *Vibrio alginolyticus* ([st306]NhaB) [45] was in agreement with the TopScreen results and the hydrophobic regions that correspond to the reentrant loops in the MemGen model were identified as periplasmic/cytoplasmic loops. However, the NhaB model contains 4 TMS in the N-terminal half, while the MemGen model contains 6 TMSs (Figure 3, Model C). Moreover, a reentrant loop was proposed between the 1st and 2nd TMS. This example shows a limitation of the TopScreen method when different models give the same cellular disposition for the three sites tested. The [st307]TRAP-T family is another example. While no experimental data is available, a consensus structure based on TMHMM [40] predictions of a number of sialic acid transporters (SiaQM) was proposed that shows 7 TMSs in the transporter domain [41]. The structure was proposed to have evolved from a protein with two homologous domains containing 5 TMSs each in which 2 TMSs were inserted in between the two domains. TopScreen does not discriminate between this model and the MemGen model that predicts a topology according to the core structure of class ST[3]. Finally, the [st303]DASS family contains many transporters of eukaryotic origin among which the well-studied human Na^+-dependent dicarboxylate transporters NaDC1 and NaDC3 found in the apical and basolateral membranes of cells of renal proximal tubule, respectively. Topology models of these transporters are based on hydropathy profile analyses and secondary structure prediction algorithms and reveal 11 to 14 TMSs [46]. The models were never tested thoroughly even though a significant body of SCAM (substituted-cysteine accessibility method) data is available. The topology model in use for the best studied transporter in the family NaDC1 shows 11 TMS with the N- and C-termini intra en extracellular, respectively, which is in line with the TopScreen data. The MemGen model adds the two reentrant loops to this model (Figure 3, Model C). An alternative model based on studies of human NaDC3 contains a reentrant loop between TMSs VIII and IX [47].

4.3. Structural diversity in transporter class ST[3]

The core structure common to the transporters in class ST[3], and also the complete structure in many of the families (Figure 3), consists of two homologous domains with inverted topology and each containing 5 TMSs with a reentrant loop in between the 4th and 5th TMS. Variations found in different families include an additional TMS at the N-terminus (1+5+5), two additional TMSs in

between the two domains (5+2+5), and an additional TMS at the C-terminus (5+5+1). Combinations of these additions were observed as well (i.e. [st325]LctP1). Some of the additional TMSs are characteristic for a particular family (structures C and G/H), while in other families their presence appears to be optional. In many families, a minority of its members have additional TMSs at the N- or C-termini. These were not included in Figure 3 because they were not considered to be typical. Similar variations in the number of TMSs are observed in transporters of other structural classes. X-ray structures of transporters from 5 different families in class ST[2] show a core structure formed by an inverted structural repeat of 2 times 5 TMSs (different from the proposed ST[3] core structure) [2-8]. In addition, two TMSs are found at the N-terminus of BetP of the betaine/choline/carnitine transporters (BCCT) family [5] and two at the C-termini of LeuT of the neurotransmitter sodium symporter (NSS) family [2], Mhp1 of the nucleobase-cation-symport-1 (NCS1) family [4] and AdiC and ApcT of the amino acid/polyamine/organocation (APC) superfamily [6-8]. In the same class, vSGLT in the sodium solute symporter (SSS) family has one additional N-terminal TMS and three additional C-terminal TMSs [3]. In class ST[1] containing the Major Facilitator Superfamily (MFS) transporters, the core structure is formed by two homologous domains containing 6 TMSs each. Transporters in the drug/H^+ antiporter family (TC 2.A.1.3 DHA) consist of 14 TMSs with the two extra TMSs inserted in between the two domains, similar to the [st325]LctP transporters in class ST[3] [48]. The TMSs that are in addition to the core structures are found at the boundaries of the core, i.e. at the N- and C-termini and in between the two repeats which leaves the structural integrity of the core intact. An insertion in the central loop is not exceptional when the core is considered to be build from two domains with the same fold that fold upon one another [17,18]. The two domains are connected by the central loop, but, from a structural point of view, might as well be connected by the N- and C-terminal ends of the core, while having the central loop disconnected. The latter structure would correspond to a swapping of the two domains as is actually observed in the [st312]NHAC family.

The inversed order of the two domains of the members of subfamilies [st312]NHAC1 and [st312]NHAC3 was identified by sequence analysis before [16] and experimentally confirmed here. From an evolutionary point of view, the presence of both orders of the two domains is easily explained but in nature both are rarely observed [49]. A remarkable feature of the proteins in subfamilies [st312]NHAC1 and [st312]NHAC3 is that the loop connecting the two domains is

located outside the cell. With few exceptions, known membrane proteins that consist of two structural repeats have the connecting loop in the cytoplasm where they may be protected from a hostile environment giving them an evolutionary advantage [50]. The Na^+/H^+ antiporter NHAC and the H^+-malate/Na^+lactate antiporter MLEN both from *B. subtilis* are the only functional characterized transporters with a swapped domain organization [34,35]. Recently, the two domains of GltS of *E. coli* from the [st324]ESS family in class ST[3] were swapped by genetic engineering [51]. $GltS^{swap}$ was equally active as the wild type protein demonstrating that from a structural or functional perspective the order of the two domains is irrelevant.

One structural variation in class ST[3] was identified that affected the 5+5 core structure. Hydropathy profile alignment showed that secondary transporters from the families [st310]Ato, [st313]AITB and two subfamilies of [st312]ArsB are missing the first TMS of the core structure, rendering a 4+5 topology (Figure 3). In case of the [st312]ArsB subfamilies, deletion of the N-terminal TMS was confirmed by sequence alignment with the other subfamilies. The data suggests that the core structure as a whole is not essential for functional and structural integrity of the class ST[3] transporters.

References

47. M.H. Saier, A functional-phylogenetic classification system for transmembrane solute transporters, *Microbiol. Mol. Rev.* **64** (2000) 354-411.
48. A. Yamashita, S.K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters, *Nature* **437** (2005) 215-223.
49. S. Faham, A. Watanabe, G. Mercado Bessemer, D. Cascio, A. Szpecht, B.A. Hirayama, E.M. Wright, J. Abramson, The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport, *Science* **321** (2008) 810-814.
50. S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E.P. Carpenter, N.G. Rutherford, J.M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S.G. Patching, R.J. Hope, H.T. Norbertczak, P.C. Roach, S. Iwata, P.J. Henderson, A.D. Cameron, Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter, *Science* **322** (2008) 709-713.
51. S. Ressl, A.C. Terwisscha van Scheltinga, C. Vonrhein, V. Ott, C. Ziegler, Molecular basis of transport and regulation in the Na+/betaine symporter BetP, *Nature* **458** (2009) 47-52.
52. Y. Fang, H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, C. Miller, Structure of a prokaryotic virtual proton pump at 3.2 Å resolution, *Nature* **460** (2009) 1040-1043.
53. X. Gao, F. Lu, L. Zhou, S. Dang, L. Sun, X. Li, J. Wang, Y. Shi, Structure and Mechanism of an Amino Acid Antiporter, *Science* **324** (2009) 1565-68.
54. P.L. Shaffer, A. Goehring, A. Shankaranarayanan, E. Gouaux, Structure and mechanism of a Na^+-independent amino acid transporter, *Science* **325** (2009) 1010-1014.
55. J.S. Lolkema, D.J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, *Mol. Membr. Biol*. **15** (1998) 33-42.
56. J.S. Lolkema, D.J. Slotboom, Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins, *FEMS Microbiol. Rev*. **22** (1998) 305-322.
57. J.S. Lolkema, D.J. Slotboom, The major amino acid transporter superfamily has a similar core structure as Na^+-galactose and Na^+-Leucine transporters, *Mol. Membr. Biol*. **25** (2008) 567-570.

58. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*, *Science* **301** (2003) 610-615.
59. Y. Huang, M.J. Lemieux, J. Song, M. Auer, D.N. Wang, Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*, *Science* **301** (2003) 616-620.
60. D. Yernool, O. Boudker, Y. Jin, E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, *Nature* **431** (2004) 811-818.
61. S. Prakash, G. Cooper, S. Singhi, M.H. Saier Jr., The ion transporter superfamily, *Biochim. Biophys. Acta*. **1618** (2003) 79-92.
62. J.S. Lolkema, D.J. Slotboom, Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis, *J. Mol. Biol*. **327** (2003) 901-909.
63. I. Sobczak, J.S. Lolkema, The 2-Hydroxycarboxylate Transporter Family: Physiology, Structure, and Mechanism, *Microbiol. Mol. Biol. Rev*. **69** (2005) 665–695.
64. A.J. Dobrowolski, I. Sobczak, J.S. Lolkema, Experimental Validation of Membrane Topology Prediction by Hydropathy Profile Alignment: Membrane Topology of the Na^+-Glutamate Transporter of *Escherichia coli*, *Biochemistry* **46** (2007) 2326-2332.
65. R. ter Horst, J.S. Lolkema, Rapid screening of membrane topology of secondary transport proteins, *Biochim. Biophys. Acta* **1798(3)** (2010) 672-80.
66. F. Baneyx, G. Georgiou, In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT, *J. Bacteriol.* **172** (1990) 491-4.
67. L.M. Guzman, D. Belin, M.J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter, *J. Bacteriol*. **177** (1995) 4121-4130.
68. E.R. Geertsma, B. Poolman, High-throughput cloning and expression in recalcitrant bacteria, *Nature Methods* **4** (2007) 705-707.
69. C. Manoil, Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions, *Methods in cell biology* **34** (1991) 61-75.
70. D. Drew, D. Sjöstrand, J. Nilsson, T. Urbig, C.N. Chin, J.W. de Gier, G. von Heijne, Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis, *Proc. Natl. Acad. Sci. USA* **99** (2002) 2690–2695.

71. B.P. Krom, J.B. Warner, W.N. Konings, J.S. Lolkema, Complementary metal ion specificity of the metal-citrate transporters CitM and CitH of *Bacillus subtilis*, *J. Bacteriol.* **182** (2000) 6374-6381.
72. C. Rensing, M. Ghosh, B.P. Rosen, Families of soft-metal-ion transporting ATPase, *J. Bacteriol.* **181** (1999) 5891-5897.
73. O.B. Kim, G. Unden, The L-Tartrate/Succinate antiporter TtdT (YgjE) of L-Tartrate fermentation in *Escherichia coli*, *J. Bacteriol.* **189(5)** (2007) 1597-1603.
74. N. Peekhaus, S. Tong, J. Reizer, M.H. Saier Jr., E. Murray, T. Conway, Characterization of a novel transporter family that includes multiple *Escherichia coli* gluconate transporters and their homologues, *FEMS Microbiol. Lett*. **147** (1997) 233-238.
75. P. Engel, R. Krämer, G. Unden, Transport of C_4-dicarboxylates by anaerobically grown *Escherichia coli:* energetics and mechanism of exchange, uptake and efflux, *Eur. J. Biochem.* **222** (1994) 605-614.
76. S. Six, S.C. Andrews, G. Unden, J.R. Guest, *Escherichia coli* possesses two homologous anaerobic C_4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct), *J. Bacteriol*. **176** (1994) 6470-6478.
77. G. Unden, J. Bongaerts, Alternative respiratory pathways of *Escherichia coli:* energetics and transcriptional regulation in response to electron acceptors, *Biochim. Biophys. Acta* **1320** (1997) 217-234.
78. E. Pinner, E. Padan, S. Schuldiner, Cloning, sequencing and expression of the NhaB gene, encoding a Na^+:H^+ antiporter in *Escherichia coli*, *J. Biol. Chem*. **267** (1992) 11064-11068.
79. E. Zientz, S. Six, G. Unden, Identification of a third secondary carrier (DcuC) for anaerobic C_4-dicarboxylate transport in *Escherichia coli:* roles of the three Dcu carriers in uptake and exchange, *J. Bacteriol*. **178** (1996) 7241-7247.
80. M. Ito, A.A. Guffanti, J. Zemsky, D.M. Ivey, T.A. Krulwich, Role of the *nhaC*-encoded Na^+/H^+ antiporter of alkaliphilic *Bacillus firmus* OF4, *J. Bacteriol*. **179** (1997) 3851-3857.
81. Y. Wei, A.A. Guffanti, M. Ito, T.A. Krulwich, *Bacillus subtilis* YqkI is a novel malic/Na+-lactate antiporter that enhances growth on malate at low protonmotive force, *J. Biol Chem*. **275(39)** (2000) 30287-30292.

82. E.L. Carter, L. Jager, L. Gardner, C.C. Hall, S. Willis, J.M. Green, *Escherichia coli abg* genes enable uptake and cleavage of the folate catabolite *p*-aminobenzoyl-glutamate, *J. Bacteriol.* **189** (2007) 3329-3334.
83. K. Abe, F. Ohnishi, K. Yagi, T. Nakajima, T. Higuchi, M. Sano, M. Machida, R.I. Sarker, P.C. Maloney, Plasmid-encoded asp operon confers a proton motive metabolic cycle catalyzed by an aspartate-alanine exchange reaction, *J. Bacteriol*. **184(11)** (2002) 2906-2913.
84. M.F. Núñez, M.T. Pellicer, J. Badia, J. Aguilar, L. Baldoma, The gene *yghK* linked to the *glc* operon of *Escherichia coli* encodes a permease for glycolate that is structurally and functionally similar to L-lactate permease, *Microbiology* **147** (2001) 1069-1077.
85. B.P. Krom, R. Aardema, J.S. Lolkema, *Bacillus subtilis* YxkJ is a secondary transporter of the 2-hydroxycarboxylate transporter family that transports L-malate and citrate, *J. Bacteriol.* **183(20)** (2001) 5862-5869.
86. A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* **305** (2001) 567-580.
87. C. Mulligan, M. Fischer, G.H. Thomas, Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archea, *FEMS Microbiol. Rev*. **35** (2010) 68-86.
88. J. Wu, L.S. Tisa, B.P. Rosen, Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase, *J. Biol. Chem*. **267** (1992) 12570-12576.
89. P. Golby, D.J. Kelly, J.R. Guest, S.C. Andrews, Topological analysis of DcuA, an anaerobic C4-dicarboxylate transporter of *Escherichia coli*, *J. Bacteriol*. **180(18)** (1998) 4821-4827.
90. K. Nanatani, T. Fujiki, K. Kanou, M. Takeda-Shitaka, H. Umeyama, L. Ye, X. Wang, T. Nakajima, T. Uchida, P.C. Maloney, K. Abe, Topology of AspT, the Aspartate:Alanine Antiporter of *Tetragenococcus halophilus*, Determined by Site-Directed Fluorescence Labeling, *J. Bacteriol*. **189(19)** (2007) 7089-7097.
91. H. Enomoto, T. Unemoto, M. Nishibuchi, E. Padan, T. Nakamura, Topological study of *Vibrio alginolyticus* NhaB Na^+/H^+ antiporter using gene fusions in *Escherichia coli* cells, *Biochim. Biophys. Acta - Biomembranes* **1370(1)** (1997) 77-86.

92. A.M. Pajor, Molecular properties of the SLC13 family of dicarboxylate and sulfate transporters*, Pflugers Arch*. **451(5)** (2006) 597-605.
93. X. Bai, X. Chen, A. Sun, Z. Feng, K. Hou, B. Fu, Membrane topology structure of human high-affinity, sodium-dependent dicarboxylate transporter, FASEB Journal **21(10)** (2007) 2409-2417.
94. J. Jin, A.A. Guffanti, C. Beck, T.A. Krulwich, Twelve-transmembrane-segment (TMS) version (*TMS VII-VIII) of the 14-TMS Tel(L) antibiotic resistance protein retains monovalent cation transport modes but lacks tetracycline efflux capacity, *J. Bacteriol.* **183** (2001) 2667-2671.
95. J.S. Lolkema, A. Dobrowolski, D.J. Slotboom, Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family, *J. Mol. Biol.* **378** (2008) 596-606.
96. M. Rapp, E. Granseth, S. Seppälä, G. von Heijne, Identification and evolution of dual-topology membrane proteins, *Nature Structural & Molecular Biology* **13** (2006) 112 – 116.
97. A. Dobrowolski, J.S. Lolkema, Evolution of Antiparallel Two-Domain Membrane Proteins. Swapping Domains in the Glutamate Transporter GltS, *Biochemistry* **49** (2010) 5972–5974.

Chapter 4

The L-Tartrate/Succinate Exchanger TtdT of *Escherichia coli*

Ramon ter Horst and Juke S. Lolkema

The C_4-dicarboxylate transporters from the DASS, Dcu, DcuC and TRAP-T families, are found in the structural class ST[3] of the MemGen structural classification together with the ESS and 2HCT families. MemGen groups families of membrane proteins in structural classes by hydropathy profile analysis. Previously, transporters from 19 different families in class ST[3] were evaluated with TopScreen, an experimental topology screening method, to verify the structural classification by MemGen. It was concluded that the structural model available for transporters of the ESS and 2HCT families is also valid for the other families in class ST[3]. Here, the topological organization of the tartrate transporter TtdT from *Escherichia coli,* a member of the DASS family, was analyzed by in-frame translational fusions between a series of progressively truncated forms of the *ttdT* gene and downstream reporter genes (*gfp*, and *phoA*), and by cysteine accessibility studies. While the cysteine accessibility studies were inconclusive, the fusion data presented here showed that the core structure of the ST[3] model is adopted by TtdT. The L-tartrate:succinate exchange activity of the cloned TtdT transporter was verified, and the correlation between TtdT expression levels and transport activity was analyzed.

1. Introduction

1.1. Evolutionary origin of C_4-dicarboxylate transporters

MemGen structural class ST[3] groups 32 families of membrane proteins, mostly from bacterial origin, that are believed to share a similar fold in the membrane [1]. Characterized members in the different families are secondary transporters, specific for organic and inorganic anions and Na^+/H^+ exchangers. Six families contain members that were shown to be C_4-dicarboxylate transporters. These are the Divalent Anion:Na^+ Symporter family ([st304]DASS), the C_4-Dicarboxylate Uptake family ([st305]Dcu), the Tripartite ATP-independent Periplasmic Transporter family ([st307]TRAP-T), the C_4-Dicarboxylate Uptake C family ([st309]DcuC, the Aspartate Alanine Exchange family ([st315]AAE) and the 2-hydroxycarboxylate transporter family ([st326]2HCT). The MemGen classification is based on structural similarity between families of membrane proteins, which is detected by hydropathy profile alignment [1]. During evolution, the structure of proteins is better conserved than the amino acid sequence and, consequently, MemGen identifies evolutionary relationships more distant than may be detected by amino acid sequence analyses. Therefore, classification in the same MemGen structural class suggests a common ancestor for the different families and divergent evolution beyond the point of significant sequence similarity. The C_4-dicarboxylate transporters in class ST[3] appear to be involved in anaerobic/fermentative metabolic pathways in bacteria, while DctA, the main C_4-dicarboxylate transporter functioning under aerobic conditions, belongs to the Dicarboxylate/Amino Acid:Cation Symporter family ([st401]DAACS) that is found in a different structural class, ST[4]. It appears that C_4-dicarboxylate transporters have evolved in parallel from different ancestors under aerobic and anaerobic conditions.

1.2. C_4-dicarboxylate metabolism

C_4-dicarboxylates like succinate, fumarate, malate, tartrate and the C_4-dicarboxylic amino acid aspartate are metabolized by bacteria under aerobic and anaerobic conditions. During aerobic growth only *uptake* systems for C_4-dicarboxylates are required, since the substrates are oxidized

to CO_2 that most likely leaves the cell by passive diffusion across the cell membrane [2]. The C_4-dicarboxylate transporter DctA that is synthesized only during aerobic growth, mediates uptake of dicarboxylates in symport with protons [3,4]. Under oxic conditions, DctA is the major C_4-dicarboxylate transporter. A *dctA* deletion mutant of *Escherichia coli* showed only poor growth on C_4-dicarboxylates [5]. Similarly, a *dctA* knock out mutant of *Bacillus subtilis* cannot grow aerobically on succinate or fumarate as the sole carbon source [6]. Aerobic degradation of C_4-dicarboxylates to CO_2 proceeds via the C_4 part of the citric acid cycle where succinate is converted to oxaloacetate via fumarate and malate. Acetyl-CoA is generated by oxidative decarboxylation of malate yielding pyruvate followed by the decarboxylation of the latter by pyruvate dehydrogenase.

Under anaerobic conditions the same pathway may run in reverse from oxaloacetate to succinate (reductive mode) [2,7,8]. Redox equivalents have to be derived from an oxidizable substrate. Oxaloacetate is reduced to malate by malate dehydrogenase. Malate is dehydrated by fumarase yielding fumarate that is reduced by fumarate reductase to succinate. Succinate is the final product and is not metabolized further but rather excreted from the cell. Substrates like aspartate, L-tartrate and also the tricarboxylate citrate may enter the pathway at different levels. Aspartate is converted to fumarate by aspartase [9-11] and L-tartrate and citrate are converted to oxaloacetate by citrate lyase and L-tartrate dehydratase, respectively (see Fig. 1) [12-14].

During co-metabolism with reduced carbon sources like sugars, externally added fumarate is utilized as an electron acceptor for fumarate respiration and, after uptake reduced to succinate by fumarate reductase. Alternatively, in many fermentation reactions succinate is formed from sugars [15-17]. The succinate formed during fumarate respiration or sugar fermentation cannot be further metabolized under anaerobic conditions and is excreted in exchange for fumarate, malate, or aspartate [4,18-20]. Thus, during anaerobic growth, C_4-dicarboxylate transporters catalyze *exchange*, *efflux* or *uptake*. The DcuA and DcuB transporters are operating during fumarate respiration [2,21,22]. They catalyze mainly fumarate:succinate antiport, but if no counter substrate is present, uptake and efflux of C_4-dicarboxylates are catalyzed as well [22,23]. DcuB is an exchanger with broad substrate specificity that catalyzes heterologous exchange of succinate and fumarate, malate or aspartate [22,23]. DcuB can be replaced by the DcuA and DcuC (mainly efflux) transporters. DcuB is synthesized during anaerobic growth during fumarate respiration, the *dcuA* gene is constitutively expressed and DcuC is produced during glucose fermentation and

catalyzes succinate efflux in this fermentation pathway [24-28]. L-Tartrate appears to be the only physiological C_4-dicarboxylate of *E. coli* that is not transported by the DcuA, B, or C carriers. L-Tartrate requires TtdT for transport [14]. TtdT catalyzes L-tartrate/succinate exchange and operates preferentially in the direction of L-tartrate uptake and succinate export. In *E. coli*, transport of the tricarboxylate citrate is mediated by CitT, that catalyzes exchange of citrate and succinate.

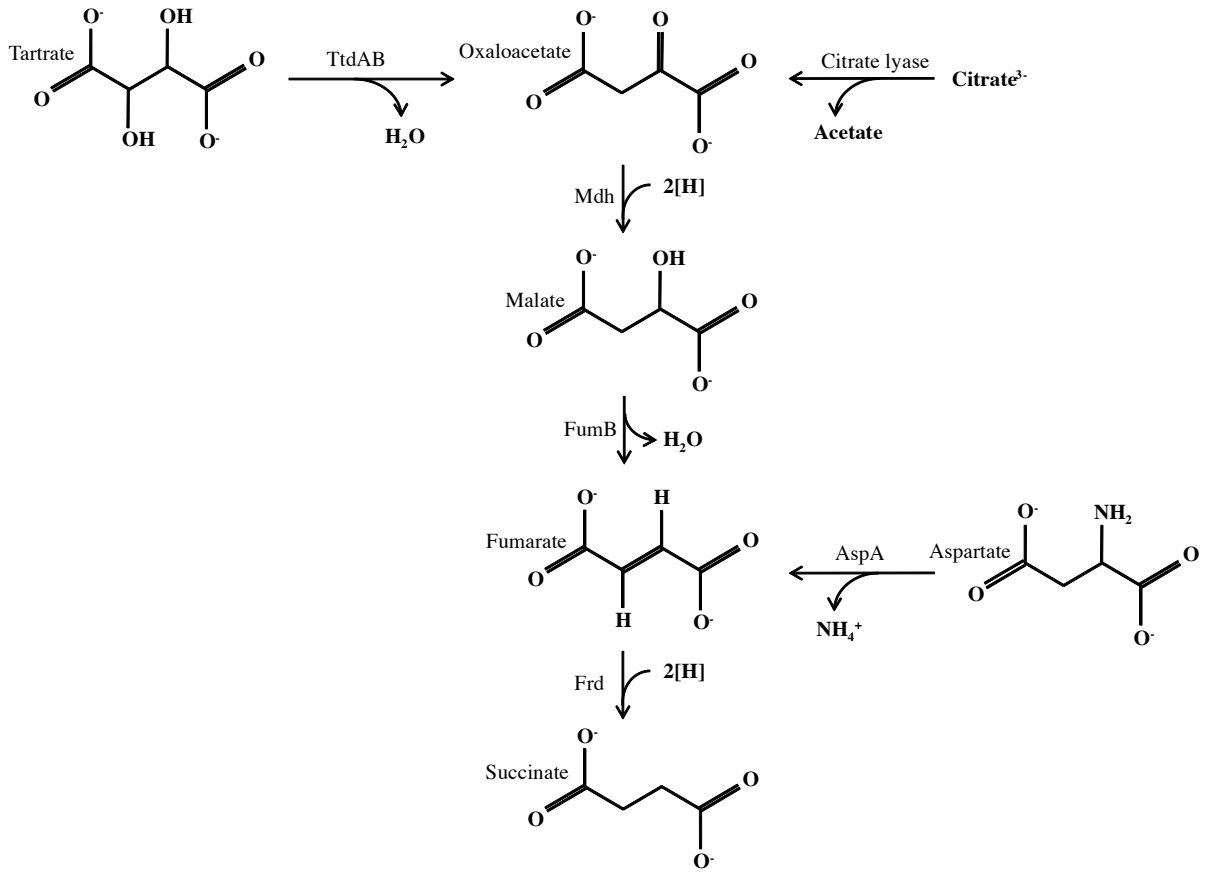


Figure 1. Pathways for (anaerobic) conversion of l-tartrate, oxaloacetate, malate, fumarate, aspartate and citrate in *Escherichia coli*. The pathways require a supply of external electron donors for fumarate and oxaloacetate reduction. The enzymes involved are: L-tartrate dehydratase (TtdAB), malate dehydrogenase (Mdh), fumarase (FumB), fumarate reductase (Frd), aspartase (AspA), and citrate lyase (CL).

An alternative pathway for the fermentation of C_4-dicarboxylates is used by a large number of bacteria that decarboxylate the substrate yielding a C_3-monocarboxylate, i.e. oxaloacetate is converted to pyruvate, malate to lactate or pyruvate, succinate to propionate, and aspartate to

alanine. Except for the decaboxylation of malate to pyruvate, these pathways are part of a much larger group of degradative decarboxylation routes that function in proton motive force generation and acid stress resistance. The autonomous pathways consists of a precursor:product exchanger that takes up the substrate from the medium and, at the same time, excretes the decarboxylation product. Proton motive force is generated in two separate steps. Membrane potential is generated because of a charge difference between the exchanged substrates of the transporter and a pH gradient follows from the consumption of a proton in the decarboxylation reaction in the cytoplasm (reviewed in [29]). Well known example of such a pathway is the malolactic fermentation pathway in lactic acid bacteria by which malate is converted to lactate which involves the malate:lactate exchanger MleP [29,30]. Oxaloacetate decarboxylation is an intermediate step in citrate fermentation, which is a similar but more complicated pathway that requires co-metabolism with a sugar [31,32]. The transporter involved is CitP that exchanges citrate and lactate. In *Streptococcus bovis,* malate is taken up by MalP in symport with protons. Malate is oxidatively decarboxylated to pyruvate yielding NADH [32,33].

1.3. *C_4-dicarboxylate transporter families*

ST[3]/[st303]DASS. The [st303]DASS family, also known as SLC13, consists of two structural subfamilies, both of which have representatives in all three domains of life, Archaea, Bacteria and Eukarya. Bacterial members are most frequently observed in current databases. Substrates of characterized members of the family are organic di- and tricarboxylate Krebs cycle intermediates as well as dicarboxylate amino acids, and inorganic sulfate and phosphate ions. Characterized transporters in this family catalyze Na^+-dependent symport [35] and heterologous exchange [36]. The L-tartrate:succinate exchanger TtdT and the citrate:succinate exchanger CitT, involved in the fermentative breakdown of L-tartrate and citrate to succinate, respectively, are members of the DASS family [37]. TtdT and CitT of *E. coli* share 40 % sequence identity. Characterized eukaryotic members are the SOT1 protein that is located in the inner envelope membrane of spinach chloroplasts, where it catalyzes the import of 2-oxoglutarate in exchange for stromal malate [38]. Other eukaryotic transporters in the family are the human Na^+-dependent dicarboxylate transporters NaDC1 and NaDC3 found in the apical and basolateral membranes of cells of renal proximal tubule, respectively.

ST[3]/[st305]Dcu. All members of the [st305]Dcu family are found in the bacterial domain. The C_4-dicarboxylate transporters DcuA and DcuB in the Dcu family are found in facultative anaerobic bacteria capable of fumarate respiration. Preferentially, DcuA and DcuB, operate in the exchangers or uptake modes of transport. C_4-dicarboxylate exchange is electro-neutral, and uptake is observed in the absence of internal substrates [21-23]. Reported topology models for DcuA [39] and DcuB [40] of *E. coli.* reveal two diverse structures, with two times 5 TMSs and 12 TMSs, respectively. Both models presented the N- and C-termini outside the cell and a cytoplasmic central loop, which is in agreement with the MemGen model and TopScreen validation study [39]. The localization data determined by DcuA-BlaM fusions, DcuB-LacZ/PhoA fusions and accessibility studies do not discriminate between the MemGen predictions (Fig. 2) and these models [39,40].

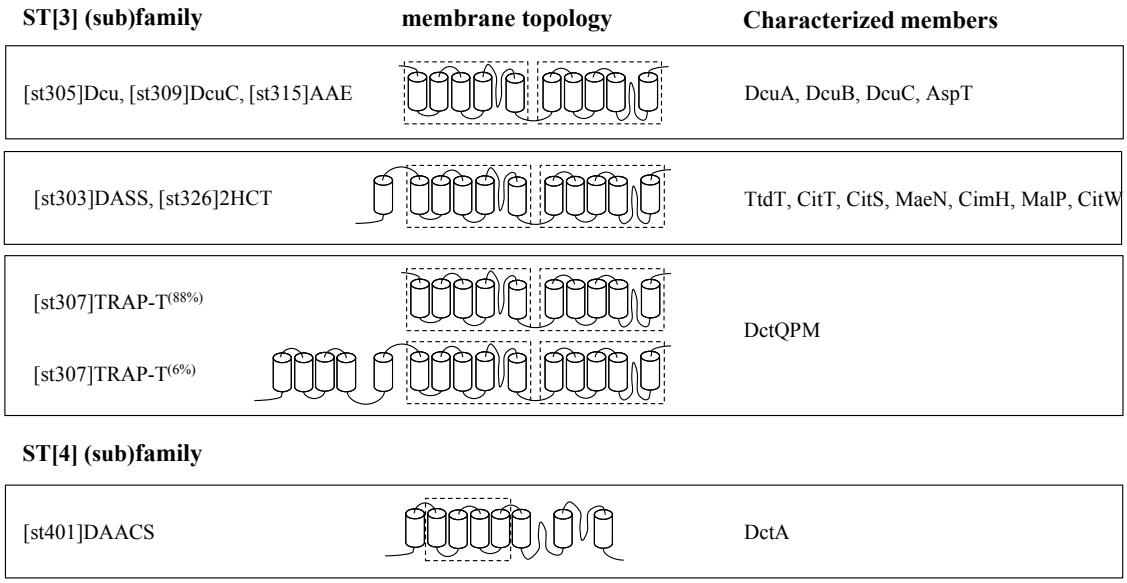


Fig. 2. C_4-dicarboxylate transporters. Secondary transporters specific for C_4-dicarboxylate transport, their structural classification and membrane topology predicted by MemGen. See text for further details.

ST[3]/[st307]TRAP-T. Transporters of the [st307]TRAP-T family are abundantly found in the bacterial and archaeal domains but not in eukaryotes. They are exceptional, as they are believed to be part of a complex that forms a transport system that appears to be a hybrid of a secondary and an ABC type of transport system [41]. The complex consists of a transporter subunit M, a smaller integral membrane protein subunit Q containing 4 TMSs and a periplasmic binding

protein P [41]. In a small fraction of the transporters (~6%), the small subunit Q is fused to the N-terminus of the M subunit via a linker TMS [42] (Fig. 2). Even smaller fractions of the M subunits contain insertions of a pair of TMSs in between the two domains or between the second and third TMS of the C-domain (not indicated). The TRAP transporters recognize a range of organic anions that are transported into the cell for further metabolism. The DctPQM complex from *Rhodobacter capsulatus* has been characterized at the functional and genetic level, and catalyzes the uptake of C_4-dicarboxylates as carbon and energy sources [43,44].

ST[3]/[st309]DcuC. Members of the [st309]DcuC family consists of two subfamilies. It is a typical bacterial family of transporters that operate in C_4-dicarboxylate metabolism in facultatively anaerobic bacteria, and are similar in function to transporters of the [st305]Dcu family. The proteins DcuC and DcuD of *E. coli* are the only members that were studied [24,26]. Only DcuC was functionally characterized. DcuC is capable of the same transport acitivities as DcuA and DcuB, but the main function is succinate efflux [21,26]. The MemGen topology model is the only membrane topology model of [st309]DcuC transporters that has been experimentally validated (by TopScreen [42,45]; Fig. 2).

ST[3]/[st315]AAE. Transporters that belong to the [st315]AAE family are conserved in many microorganisms throughout the bacterial domain [46,47]. A functionally characterized protein that belongs to this family is AspT of the Gram-positive lactic acid bacterium *Tetragenococcus halophilus*. AspT catalyzes aspartate:alanine exchange as part of a degradative amino acid decarboxylation pathway. Following uptake into the cell, the C_4 amino acid aspartate is decarboxylated to yield alanine. The transporters in the family are characterized by a large hydrophilic loop in the middle of the protein that is likely to fold into a membrane associated domain. AspT was shown to contain 10 TMSs and periplasmic N- and C-termini [48,49], which is in agreement with the MemGen model and TopScreen validation study [42]. The large hydrophilic loop is located between TMSs 5 and 6, in between the two putative transmembrane domains. ST[3]/[st326]2HCT. Members from the [st326]2HCT family are found exclusively in the bacterial kingdom. Characterized proteins are Na^+ or H^+ symporters, or precursor/product exchangers involved in generation of secondary metabolic energy. Substrates of characterized members of the [st326]2HCT family are citrate or malate. Characterized Na^+ or H^+ symporters are the citrate transporter CitS of *Klebsiella pneumoniae* and the malate transporter MaeN of Bacillus subtilis [50-52]. The citrate/malate transporter CimH of *B. subtilis* and the malate

transporter MalP of *Streptococcus bovis* were shown to be H^+ symporters [34,53]. MleP of *Lactococcus lactis,* CitP of *Leuconostoc mesenteroides* and CitW of *Klebsiella pneumoniae* catalyze malate:lactate, citrate:lactate and citrate:acetate exchange, respectively [54,55,30]. The Na^+-citrate transporter CitS of *Klebsiella pneumoniae* is the best studied transporter in the [st326]2HCT family, and its structural model will be discussed in more detail below.

ST[4]/[st401]DAACS. DctA transporters belong to a large family of secondary transporters, the Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) family, that also comprises well-characterized glutamate/aspartate transporters and neutral amino acid transporters [37]. They are represented in Archaea, Bacteria and Eukaryotes. Transporters of the bacterial DctA family catalyze H+ or Na+ symport with C_4-dicarboxylates. Detailed studies of the glutamate transporters GltT of *Bacillus stearothermophilus* revealed a membrane topology of 8 TMS with two reentrant loops in the C-terminal half of the proteins, and a cytoplasmic N- and C-terminus (Fig. 2) [56-58]]. The high-resolution structure of the Na^+-aspartate transporter Glt_{Ph} [59] supported these findings.

1.4. *Structure of class ST[3] C_4-dicarboxylate transporters*

Together with the ESS and 2HCT families, the DASS, Dcu, DcuC and TRAP-T families are found in structural class ST[3] of the MemGen structural classification, where families of membrane proteins are grouped in structural classes by hydropathy profile analysis [60]. Detailed studies of the Na^+-citrate transporter CitS of *Klebsiella pneumonia* and the Na^+-glutamate transporter GltS of *Escherichia coli* [61,62], members of the [st326]2HCT and [st324]ESS families in class ST[3] have resulted in a model for the core structure of ST[3] transporters. The core structure consists of two homologous domains, each containing 5 transmembrane segments and a reentrant loop in between the 4^{th} and 5^{th} segment. The two domains have opposite orientations in the membrane [61]. Previously, transporters from 19 different families in class ST[3], including the DASS, Dcu, DcuC and TRAP-T families, were evaluated by the TopScreen experimental topology screening method to verify the structural classification by MemGen [42]. TopScreen involves the determination of the cellular disposition of three sites in the polypeptide chain of the proteins, the N- and C-terminus and a position in the central loop [45]. Comparison to the experimental data showed that in all cases the prediction by MemGen was correct, and

nearly all of the identified and experimentally confirmed structural variations involve additions of TMSs at the boundaries of the core model, at the N- and C-termini or in between the two domains. It was concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other families in class ST[3]. The topology models of the C_4-dicarboxylate transporters found in structural class ST[3] and ST[4] of the MemGen classification are shown in Fig. 2.

1.5. L-Tartrate transporter TtdT

The putative L-tartrate transporter TtdT is a member of the [st303]DASS family in the MemGen classification, suggesting a similar structure and transport mechanism as proposed for the Na^+-citrate transporter CitS and the Na^+-glutamate transporter GltS in the [st326]2HCT and [st324]ESS families, respectively. Models predicted by secondary structure predictors and presented for eukaryotic members of the DASS family are at variance with the MemGen model. Here, the activity of the TtdT protein in L-tartrate:succinate exchange was confirmed and the membrane topological organization of TtdT was analyzed by in-frame translational fusions between a series of progressively truncated forms of the *ttdT* gene and downstream reporter genes (*gfp*, and *phoA*).

2. Materials and Methods

2.1 Materials

Phusion DNA polymerase was obtained from Finnzymes (Espoo, Finland). T4 ligase was obtained from New England Biolabs (Frankfurt am Main, Germany). All other enzymes were obtained from Fermentas (Burlington, Canada). Mutagenic oligonucleotides were obtained from Sigma-Aldrich
(Zwijndrecht, Netherlands), or from Operon (Ebersberg, Germany). *p*-Nitrophenyl phosphate (pNPP) and *N*-ethylmaleimide (NEM) were obtained from Sigma (Zwijndrecht, The Netherlands), fluorescein maleimide (FM) from Invitrogen (Carlsbad, United States), ampicilin from Roche Diagnostic GmbH (Mannheim, Germany) and arabinose from Sigma-Aldrich GmbH

(Steinheim, Germany).
Radio-labeled substrates D,L-[1,4-^{14}C]-tartrate (3.7-4.44 GBq/mmol) and [1,4-^{14}C]-succinate (1.48-2.22 GBq)/mmol were obtained from Hartmann Analytic GmbH (Braunschweig, Germany).

2.2. Bacterial strains and growth conditions

Escherichia coli strain SF100 (***recA Δlac ΔompT***) [63] harboring the indicated pLIC vector (see below), was routinely grown in Luria Broth medium at 37 ^{o}C, with ampicillin added at a final concentration of 50 μg/ml. Overnight cultures were diluted 30-fold in fresh medium and when the optical density measured at 660 nm (OD_{660}) reached a value between 0.6 and 0.8, arabinose was added at a final concentration of 0.0002-0.05% (w/v) to induce protein production from the plasmids. Following growth for another 1h, cells were harvested by centrifugation operated at 4^{o}C, washed once and, subsequently, resuspended in the indicated buffer and kept on ice until use.

2.3 Ligation Independent Cloning

All genetic manipulations were as described before [45]. Briefly, a synthetic double stranded piece of DNA (the LIC cassette) was inserted downstream of the arabinose promoter in the commercial pBAD24 vector (Invitrogen) [64]. The nucleotide sequence of the sense strand was 5'-C ATG GGT *CAT CAT CAC CAT CAC CA**T*** **TTA AA**T AGT GGT GTG GTA C- 3' in which the initiation codon was underlined, the sequence coding for a His_6-tag set in italics and a *SwaI* restriction site in bold. The resulting vector with a gene encoding *phoA* or *gfp* inserted in frame with the initiation codon was used to construct alkaline phosphatase or green fluorescence protein fusion proteins, and was termed pLIC1 or pLIC2 respectively. Replacing the *phoA* fragment by a double stop codon resulted in vector pLIC3, which is used to produce His-tagged proteins. Ligation Independent Cloning (LIC) was done essentially as described [65]. The LIC vectors pLIC1-3 were linearized by *Swa*I digestion. The transporter gene or gene fragments were amplified using forward and backward primers containing 5'-flanking regions corresponding to the nucleotide sequences upstream and downstream of the *Swa*I site in the LIC cassette as

follows: 5'-CATGGGTCATCATCACCATCACCATTTG.....-3' and 5'-TACCACACCACTATTTTG.....-3', respectively. One forward and five reverse primers were used to amplify a series of five progressively truncated forms of the *ttdT* gene. Single-stranded overhangs of the PCR products and vectors were generated using T4 DNA polymerase in the presence of dGTP (vector) and dCTP (PCR product). The complementary overhangs of PCR product and vector annealed upon mixing, after which the resulting heteroduplexes were transformed to *E. coli* SF100.

2.4. GFP and PhoA assays

2.4.1. PhoA assay

Cells from 0.5 ml of culture of ***E. coli*** SF100 cells harboring a pLIC1 vector carrying the indicated insert were washed once, and subsequently resuspended in 1 M Tris-HCl pH 8.0 buffer after which the OD_{660} was measured. Following incubation for 5 min at 37 ^{o}C, 200 µL of a 1.5 mM *p*-nitrophenyl phosphate (pNPP) solution was added and the suspension was incubated until a yellow color developed. Then, the cells were spun down and the absorption was read at 420 and 550 nm using a Hitachi U-1100 spectrophotometer. The phosphatase activity was calculated in Miller units [66].

2.4.2. GFP assay

Cells of ***E. coli*** SF100 harboring a pLIC2 vector carrying the indicated insert were washed once and, subsequently, resuspended in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 15 mM EDTA to a final OD_{660} of 0.2. Following incubation for 30 min at room temperature, a 150 µL cell suspension was transferred into a precision cell (Hellma, Quartz SUPRSIL) and GFP fluorescence emission intensity at 508 nm was measured at an excitation wavelength of 470 nm using an AMINCO Bowman Series 2 Luminescence Spectrometer. Background signals caused by light scattering were estimated from cells harboring the pLIC3 vector without insert [67].

2.4.3. Evaluation of data

Mean values and standard deviations were calculated from at least three independent measurements. PhoA and GFP activities of the cells were normalized by the mean PhoA activity of all positive PhoA fusions (>100 Miller units) and all positive GFP fusions (> 0.5 emission units), respectively. The logarithm of the ratio of the normalized PhoA and GFP activities was calculated for each full-length and half protein to obtain a measure for the cellular localization of the fusion point. A positive value corresponds to a periplasmic localization, a negative bar to a cytoplasmic localization. The length of the bar indicates the significance of the localization and would be independent on the expression levels [43]. Values >2.5 or <-2.5 were arbitrarily set to 2.5 and -2.5, respectively.

2.5. Transport assays

At t = 0, D,L-[1,4-^{14}C]-tartaric acid, or [1,4-^{14}C]-succinate (with or without 1mM 'cold' L-tartrate) was added to a final concentration of 9.2 μM and 20.8 μM of labeled substrate, respectively. Uptake was stopped at 0, 15, 30, 60, 120 and 240 sec by the addition of 2 ml ice-cold 0.1 M LiCl followed by immediate filtering over cellulose nitrate filter (0.45 μm pore size) and washing of the filter with an additional 2 ml of the LiCl solution. Filters were collected and their radioactivity was measured in a liquid scintillation counter. The amount of D,L-[1,4-^{14}C]-tartaric acid, or [1,4-^{14}C]-succinate taken up was calculated, assuming an OD_{578} of 1 corresponds to 281 mg dry weight per liter of *E. coli* cells [24].

2.6. Treatment with thiol reagents

E. coli SF100 cells harboring pLIC3 vectors carrying the TtdT encoding gene were washed once and, subsequently, resuspended in ice-cold 50 mM potassium phosphate buffer pH 7.0. Cells from 200 ml of culture were resuspended in 1 ml of buffer. A 5 mM solution of fluorescein 5-maleimide (FM) was freshly prepared by diluting a 50 mM solution in DMSO with 50 mM potassium phosphate buffer pH 7.0. FM was added to the cell suspension to a final concentration of 0.1 mM. After incubation for 20 min at room temperature, excess FM was quenched by adding 2 mM dithiothreitol (DTT). Cells were washed three times with 50 mM potassium phosphate buffer pH 7.0, and disrupted by sonication. Following removal of debris, membranes were

collected from the supernatant by centrifugation for 25 min at 80,000 rpm in a Beckman TLA100.2 rotor at 4 ^{o}C. The pellet was solubilized in 1 mL 50 mM potassium phosphate buffer pH 8.0 containing 400 mM NaCl, 20% glycerol, 10 mM imidazole, 1mM DTT and 2% Triton X-100 and left on ice for 30 min after which undissolved material was removed by centrifugation at 90,000 rpm for 25 min at 4 ^{o}C. The supernatant was mixed with Ni^{2+}-NTA resin (20 µL bed volume) equilibrated in 50 mM potassium phosphate buffer pH 8.0 containing 600 mM NaCl, 20% glycerol, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100, and incubated overnight at 4^{o}C under continuous shaking. The resin was pelleted by brief centrifugation in a table top centrifuge and the supernatant was removed. The resin was washed with 0.5 mL of equilibration buffer containing 300 mM NaCl and 20-40 mM imidazole. The protein was eluted with 25 µL of the same buffer at pH 7.0 and containing 200 mM imidazole. Subsequently, samples were loaded onto a 12% sodium dodecyl sulfate-polyacrylamide gel. Following electrophoresis (SDS-PAGE), in-gel fluorescence was recorded using a Fujifilm LAS-4000 luminescent image analyzer, and the gel was stained with Coomassie Brilliant Blue (CBB).

Alternatively, a 100 mM solution of *N*-ethylmaleimide (NEM) was freshly prepared in 50 mM potassium phosphate buffer pH 7.0. NEM was added to the cell suspension to a final concentration of 1 mM. After incubation for 15 min at room temperature, excess NEM was quenched by adding 2 mM DTT. Cells were washed times with 50 mM potassium phosphate buffer pH 7.0, and subsequently diluted to a final OD_{660} of 4.0 for immediate use in transport studies.

3. Results

3.1. Structural model of TtdT

The L-tartrate/succinate exchanger TtdT of *Escherichia coli* is a member of the DASS family in the Transport Classification system (TC system) [37]. Together with the ESS and 2HCT families, the DASS family is found in structural class ST[3] of the MemGen structural classification. Detailed structural models of transporters in the former two families are available [61,62]. Both models share a 'core' structure [42] with two domains of 5 TMSs each (5+5 topology) that are oppositely oriented in the membrane (inverted topology). In between the 4th and 5th segment in

each domain a reentrant loop is found. In case of the 2HCT transporters, the first domain is preceded by a single N-terminal segment. The DASS family consists of two subfamilies representing two clusters in a phylogenetic tree with significant overall sequence identity. Optimal alignment of the family averaged hydropathy profiles of the 2HCT families and the profiles of the two DASS subfamilies resulted in similarity scores [68,69] of 1.073 and 0.58, respectively, indicating very similar profiles (Figure 3). The alignments allow for projection of the 2HCT structural model (Figure 3, top) on the DASS family.

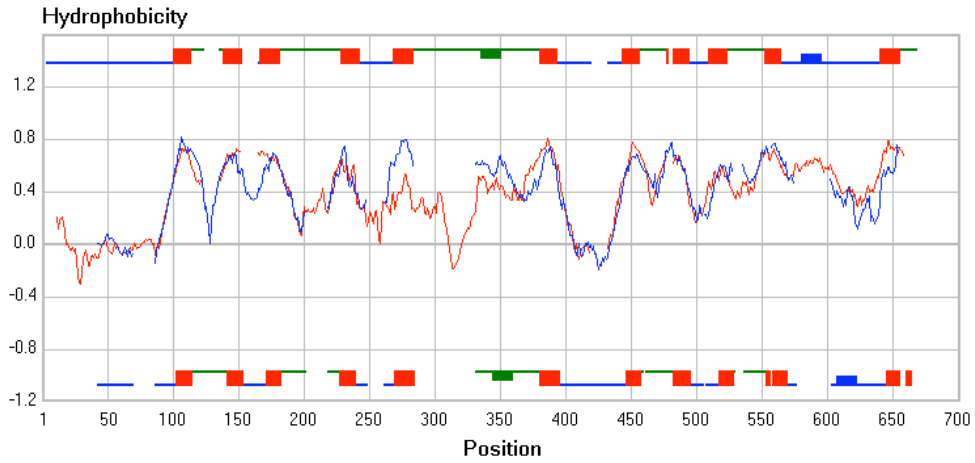


Fig. 3. Family averaged hydropathy profile alignments. Alignment of the family averaged hydropathy profiles of [st303]DASS1 (red) and (A) [st326]2HCT (blue). Topology models are depicted above ([st303]DASS), or below ([st326]2HCT) the corresponding alignments. Red squares, TMSs; green lines, external loops; blue lines, cytoplasmic loops; thickened loop regions (blue or green rectangles), reentrant loops. The Structure Divergence Score (SDS), which is a measure of the similarity of the profiles after alignment [9], was 1.073.

Similarly as observed for the 2HCT transporters, the DASS transporters contain an additional TMS preceding the core structure at the N-terminus and resulting in a total of 11 TMSs (Figure 3, bottom topology model). Structural models obtained in this way were experimentally validated for a number of ST[3] families, including the DASS family ([st303]DASS), by the TopScreen reporter fusion approach [42,45]. TopScreen involves the determination of the cellular disposition of three positions of a membrane protein, the N- and C-terminus and a position in the central loop that is measured by the localization of the C-terminus of the half-protein. Here, by using the TopScreen approach, a series of five progressively truncated forms of the *ttdT* gene was cloned into the appropriate vectors resulting in PhoA and GFP fusion proteins. The reporters were fused

behind residues D306, D330, S361, S396, and P455, which are all located in putative loop regions. Together with the previously constructed fusions to the half protein (L276) and the full-length protein [42], the set covers the C-terminal half of TdtT. Alkaline phosphatase activity of cells expressing PhoA fusions was plotted against the GFP fluorescence of cells expressing the corresponding GFP fusion and the previously reported data [42] was included for completeness (Fig. 4A). The data points formed two well-separated groups falling along the two axes. High alkaline phosphatase activity and low GFP fluorescence corresponds to a periplasmic localization of the fusion point whereas low alkaline phosphatase activity and high GFP fluorescence corresponds to a cytoplasmic localization. The normalized ratios of the PhoA and GFP signals of the individual proteins (Fig. 4B; See also Methods 2.4.3.) show that two truncated proteins (D306, S361) and the full-length protein would have their C-terminus in the periplasm and the remaining three truncates (330D, 396S, 455P) and the half protein would have their C-terminus in the cytoplasm. Together with the cytoplasmic N-terminus [42], it is concluded that the experimental data agrees with the structural model predicted by MemGen (Fig. 4C). It should be noted that, according to the experimental data (in Fig. 4), the hydrophobic region between fusion points S396 and P455 is not transmembrane and represents the reentrant loop structure in between the 4^{th} and the 5^{th} TMS in the model.

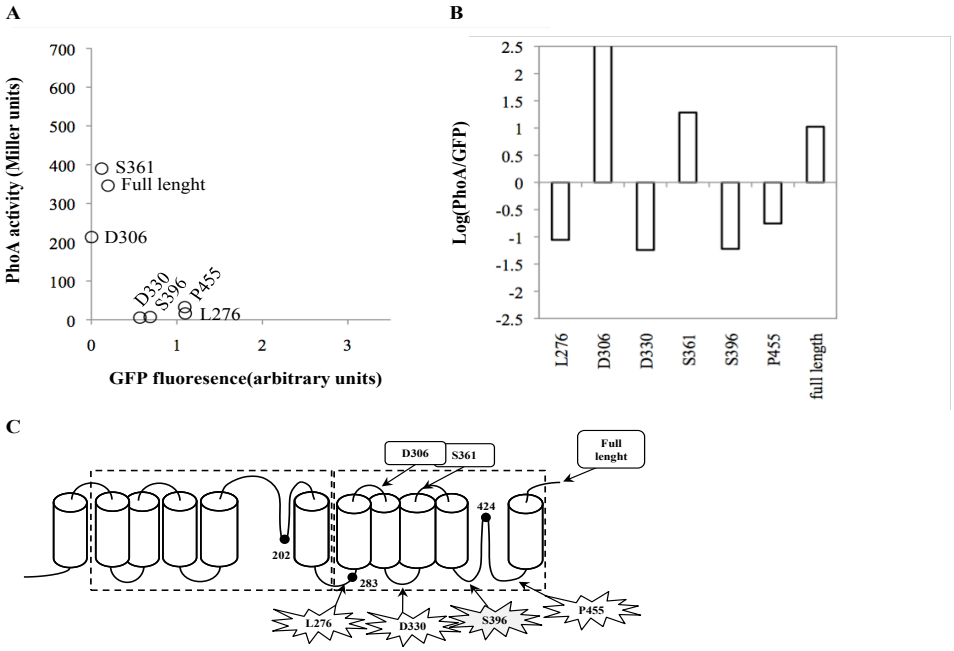


Fig. 4. C-terminal localization of TtdT truncates. PhoA activity *versus* GFP fluorescence intensities (A) and logarithm of the ratio of the normalized PhoA and GFP signals (B) for truncated proteins of TtdT of *E. coli.* The C-terminal residue of each truncated protein was indicated. (C) MemGen topology model of the [st303]DASS family. Arrows indicate the positions and results (in/out) of the experimental C-terminal localizations. Here, truncated transporter-reporter fusion proteins with GFP or PhoA activity, are represented by stars and rounded rectangles, respectively. Previously reported data has been included for completeness [42].

3.2. Transport activity of TtdT

The *ttdT* gene encoding the L-tartrate/succinate exchanger was cloned in the pLIC3 vector yielding pLIC3-TtdT, which encodes the TtdT protein extended with an N-terminal His_6-tag. *E. coli* SF100 harboring pLIC3-TtdT was grown under aerobic conditions to minimalize expression of anaerobic C_4-dicarboxylate transporters [4,18,19,21-23]. Transport of D,L-[^{14}C]-tartrate was analyzed using resting cells resuspended in 50 mM KPi pH 7. Within one minute, uptake of D,L-[^{14}C]-tartrate added at a concentration of 9.2 µM to the cells reached a level of 3.8 nmol/mg of cell protein that was maintained for at least the next 3 minutes, suggesting no further metabolism in the cytoplasm. The uptake activity was more than 10 times higher than observed in cells harbouring the vector (pLIC3) without the *tdtT* insert (Figure 5A). Presumably, the low

background activity was due to expression of the chromosomal copy of *ttdT* and/or other dicarboxylate transporters with (low) specificity for tartrate. Resting cells of *E. coli* SF100 cells harboring plasmid pLIC3-TtdT transiently accumulated [^{14}C]-succinate to finally reach a level of 0.5 nmol/mg of cell protein after 4 minutes that amounted to about 1/3 of the maximum accumulation level observed after 0.5 minutes (Fig. 5B). The behavior in the control cells harboring vector pLIC3 was comparable with only a slightly lower maximum accumulation and a slightly shorter duration of the overshoot (Fig. 5C). Apparently, the DctA transporter, which is known to be the major succinate transporter under aerobic growth conditions in *E. coli* [4,5], is mostly responsible for the uptake of succinate in both types of cells. Most likely, uptake of succinate into the cells is followed by conversion to other compounds that in part are excreted and in part retained the radiolabel in the cell. To demonstrate L-tartrate:succinate exchange by TtdT, the experiment was repeated in the presence of an excess of 1 mM 'cold' L-tartrate. In case of the control cells, the uptake of [^{14}C]-succinate was not significantly different (Fig. 5C). In contrast, with the TtdT producing cells a significant reduction in the overshoot was observed, in line with an exchange of external tartrate and internal [^{14}C]-succinate transported into the cell mainly by DctA (Figure 5B).

The results show that the TtdT protein catalyzes unidirectional uptake of tartrate and tartrate:succinate exchange. Unidirectional uptake of succinate could not be demonstrated unequivocally because of the high background uptake activity of the cells, presumably by the DctA transport protein.

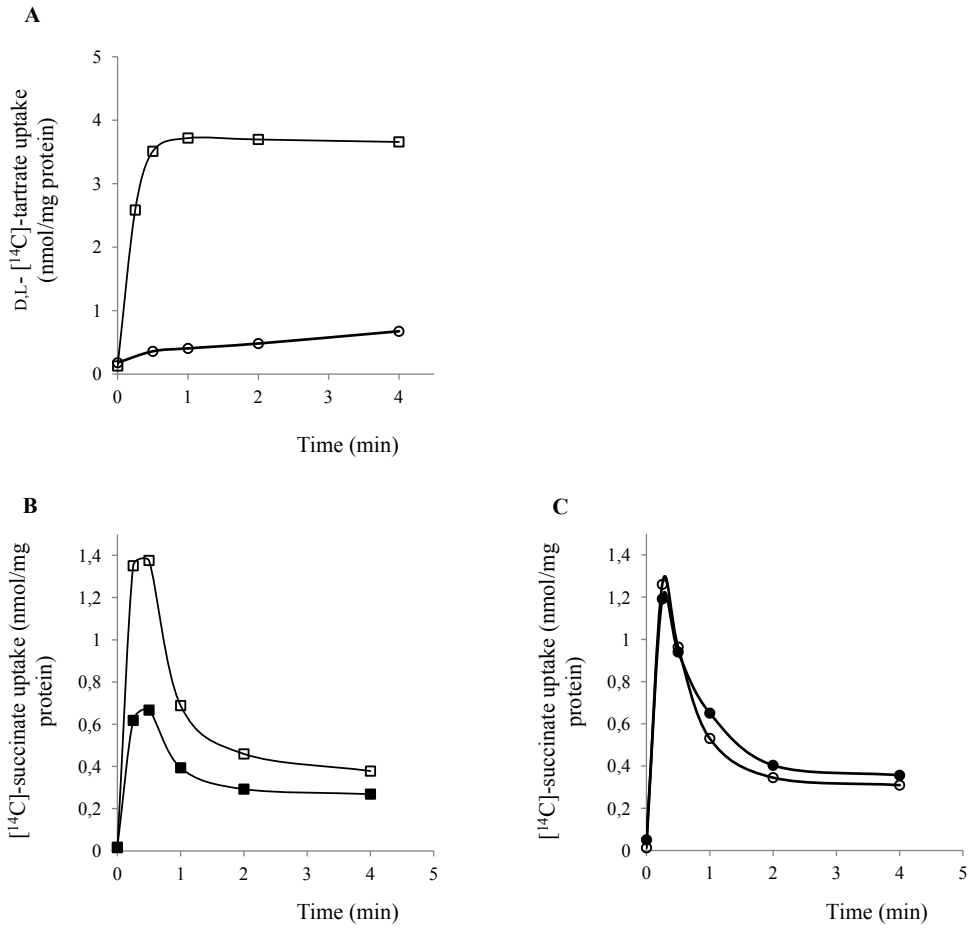


Fig. 5. Kinetics of D,L-[^{14}C]-tartrate and [^{14}C]-succinate uptake. Uptake activity in cell suspensions of *E. coli* SF100 harboring plasmid pLIC3-TtdT (□) or pLIC3 (○) plasmid with D,L-[^{14}C]-tartrate (A) or [^{14}C]-succinate (B,C) with (■ and ●) or without (□ and ○) an excess of 'cold' L-tartrate as substrates . To induce protein production from the plasmids, arabinose was added at a final concentration of $2x10^{-4}$% (w/v).

3.3. Correlation expression level and transport activity of TtdT

E. coli cells harboring plasmid pLIC3-TtdT were grown as described in the methods section in the presence of increasing concentrations of the inducer arabinose. Cells were harvested and prepared for use in transport assays and the TtdT protein was purified quantitatively to estimate the expression level in the membranes of the same cells. Prior to Ni^{2+}-NTA affinity purification of the His-tagged transporters from the membranes, the cells were treated with fluorescein maleimide to estimate labeling efficiency, as visualized by fluorescence imaging after SDS-PAGE. In the absence of inducer, no TtdT protein could be detected in the membrane, both by

Coomassie Brilliant Blue (CBB) staining of the gel or by fluorescence (Fig. 6A, CBB and F, respectively). Both signals increased between induction levels of $2x10^{-4}$ and $2x10^{-3}$ % arabinose, but when a concentration of $5x10^{-2}$ % arabinose was used a decrease in the expression level of TtdT in the membranes was observed, while the labeling efficiency still increased. The ratio of fluorescence over protein stain was highest at $5x10^{-2}$ % arabinose and more or less the same at the other two arabinose concentrations (not shown).

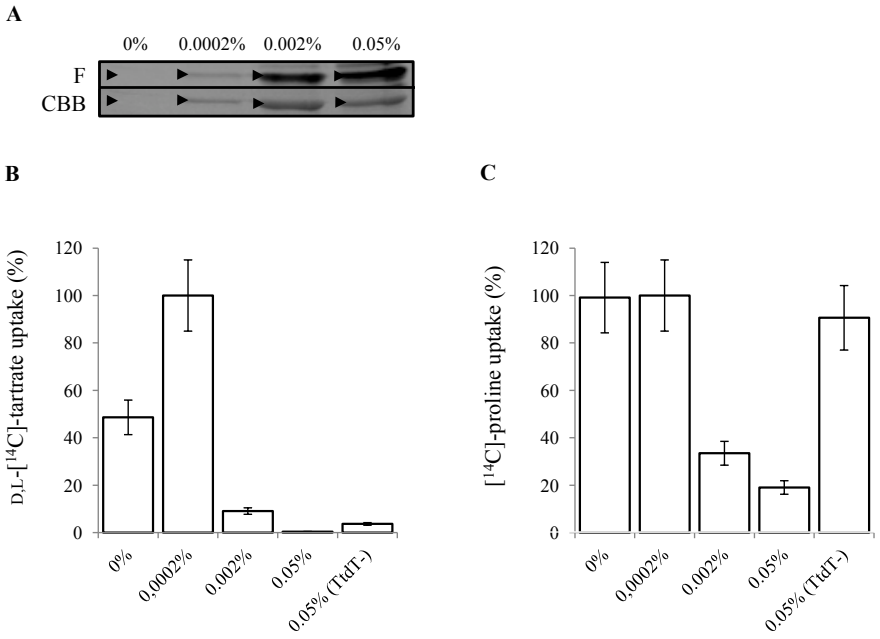


Fig. 6. Correlation expression level and transport activity of TtdT. *E. coli* SF100 cells were grown in the presence of an arabinose concentration (w/v) of 0 % (◊), 0.0002 % (□), 0.002 % (Δ) or 0.05 % (×) to induce protein expression. Cells carrying the TtdT- plasmid, induced with 0.05% arabinose, were included as a control (○). (A) Prior to Ni^{2+}-NTA affinity purification of His-tagged TtdT from the membranes, the cells were treated with fluorescein maleimide (FM) and, subsequently, visualized by fluorescence imaging after SDS-PAGE. F, fluorescence image of the part of the gel containing the TtdT protein; CBB, same part of the gel after staining with Coomassie Brilliant Blue. (B) D,L-[^{14}C]-tartrate and (C) [^{14}C]-proline uptake activity was measured relative to the uptake activity of *E. coli* SF100 carrying the TtdT plasmid induced with $2x10^{-4}$ arabinose.

D,L-[^{14}C]-Tartrate uptake activity, measured with the same batches of cells, was not proportional to the amount of TtdT protein in the membranes. Growth in the absence of inducer resulted in significantly higher uptake activities than observed in control cells containing the empty vector, demonstrating that even the tightly controlled P_{ara} promotor on the pBAD24 vector is leaky to

some extent (Fig. 6B). Highest uptake activities were observed when a concentration of $2x10^{-4}$ % arabinose was used, but the uptake activity decreased rapidly with increasing arabinose concentration, down to levels below the background activity observed in the control cells. The collaps of the uptake activity at higher TtdT protein levels in the membrane may be caused by misfolding when protein production proceeds at too high a rate or by a negative effect on the integrity of the cytoplasmic membrane caused by the over expressed TtdT protein. The arrest of growth of the cells upon induction at all arabinose concentrations was indicative of the latter option (data not shown). To further explore this possibility, [^{14}C]-proline uptake activity was measured with the same batches of cells. Uptake was at the same high level in the control cells and uninduced cells and cells induced with $2x10^{-4}$ % arabinose, but decreased rapidly when the arabinose concentration was increased to $2x10^{-3}$ and $5x10^{-2}$ % (Fig. 6C). Clearly, the over expression of TtdT is toxic to the cells. All other experiments reported here were performed with cells grown under optimal induction conditions ($2x10^{-4}$ % (w/w) arabinose), unless stated otherwise.

3.4. Sensitivity of TtdT activity to treatment with N-ethylmaleimide

Wild type TtdT contains three Cys residues, Cys202 and Cys424 in the putative reentrant loops in the N- and C-terminal domains, respectively, and Cys283 in the central cytoplasmic loop that connects the two domains (see Fig. 4). Three TtdT mutants were constructed by mutating the three native cysteine residues one by one into a serine residue yielding SCC, CSC and CCS for C202S, C283S and C424S, respectively. Transport activity of the TtdT mutants with D,L-[^{14}C]-tartrate as the substrate showed comparable uptake activities (Fig. 7A) and expression levels (Fig. 7B) for the mutants SCC, CSC and CCS as observed for the wild type TtdT, indicating that the mutations did not significantly affect functional folding of the proteins in the membrane. Treatment of SF100 cells expressing wild-type TtdT with 1 mM *N*-ethylmaleimide (NEM), a membrane permeable thiol reagent, for 15 minutes at room temperature, resulted in a 60 % lower uptake activity of D,L-[^{14}C]-tartrate (Fig. 7C). Surprisingly, treatment of the cells expressing the mutant transporters with NEM under identical conditions, as above, resulted in a similar reduction of activity for all three mutants (Fig. 7C). Apparently, not a single Cys residue in the TtdT is responsible for the loss of activity upon alkylation of the protein.

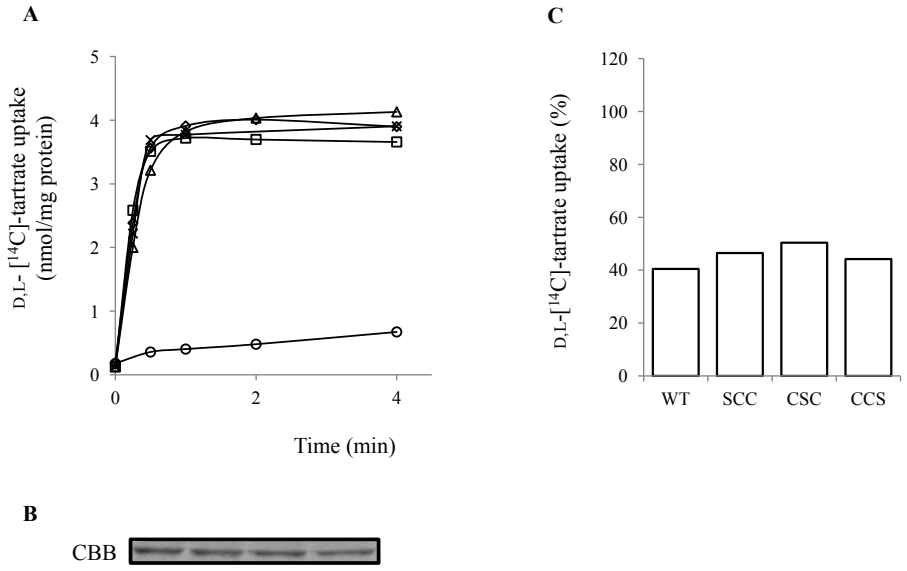


Fig. 7. D,L-[^{14}C]-tartrate uptake activity of TtdT Ser mutants. (A) D,L-[^{14}C]-tartrate uptake activity was measured in *E. coli* SF100 cells expressing TtdT Ser mutants. (B) As in (A), but prior to the uptake activity measurements, cells were treated with 1 mM NEM for 15 min, followed by the addition of 1mM DTT. The activity was plotted as the relative activity compared to the NEM-untreated cells. The amount of internal D,L-[^{14}C]-tartrate was determined by filtration assay and scintillation counting. (C) The amount of His-tagged NEM treated TtdT from the membranes was visualized by CBB staining after SDS-PAGE.

3.5. Accessibility of the native cysteine residues of TtdT from the periplasmic side of the membrane

Wild type TtdT and the SCC, CSC and CCS mutants were treated with 0.1 mM of the membrane impermeable fluorescent sulfhydryl reagent fluorescein maleimide (FM) for 20 minutes at room temperature. Control studies reported before showed that under these conditions, the thiol reagent does not enter the cell to any significant concentration and only labels Cys residues exposed to the external medium [45]. Following treatment of the cells with fluorescein maleimide and Ni^{2+}-NTA affinity purification of the His-tagged transporters from the membranes, Coomassie Brilliant Blue staining after SDS-PAGE revealed similar protein levels in the membranes (Fig. 8, panel 'CBB').

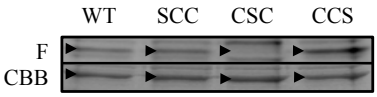


Fig. 8. Accessibility of cysteine residues in TtdT and Ser mutants of TtdT by FM. Prior to Ni^{2+}-NTA affinity purification of His-tagged protein from the membranes, cells expressing TtdT (□) or one of the Ser mutant of TtdT, SCC (◊), CSC (Δ), or CCS (×), were treated with fluorescein maleimide (FM) and subsequently visualized by fluorescence imaging after SDS-PAGE. F, fluorescence image of the part of the gel containing the protein; CBB, same part of the gel after staining with Coomassie Brilliant Blue (see Methods). Arrowheads indicate the positions of the partially purified proteins and the corresponding fluorescent band.

Fluorescence imaging of the gel showed labeling of wild type and the mutant TtdTs with the exception of CSC (Fig. 8, panel 'F'). It follows that the cysteine residue at position 283 would be the only accessible residue for labeling with FM from the periplasmic side of the membrane, a result that would be inconsistent with the location of C283 in the central cytoplasmic loop. An alternative explanation would be that (over)production of TtdT causes a cell to become leaky and permeable to FM.

4. Discussion

4.1. L-Tartrate:succinate exchange in E. coli

E. coli ferments L-tartrate under anaerobic growth conditions in the presence of an oxidizable co-substrate like glycerol or glucose. Following uptake from the medium, tartrate is converted to succinate in two subsequent dehydration and reduction steps. The first step yields oxaloacetate then malate, the second fumarate and then succinate. The pathway is largely the same as for the fermentation of citrate by *E. coli* that enters the common part of the pathway at the level of oxaloacetate by the action of citrate lyase (Fig. 1). Succinate is the end product of the pathway and is excreted into the medium. The genes encoding the tartrate and citrate specific parts of the pathway leading to oxaloacetate are clustered in operon structures on the chromosome. The genes encoding two subunits of the L-tartrate dehydratase, *ttdA* and *ttdB*, are clustered together with the putative tartrate transporter gene *ttdT* (Fig. 9A).

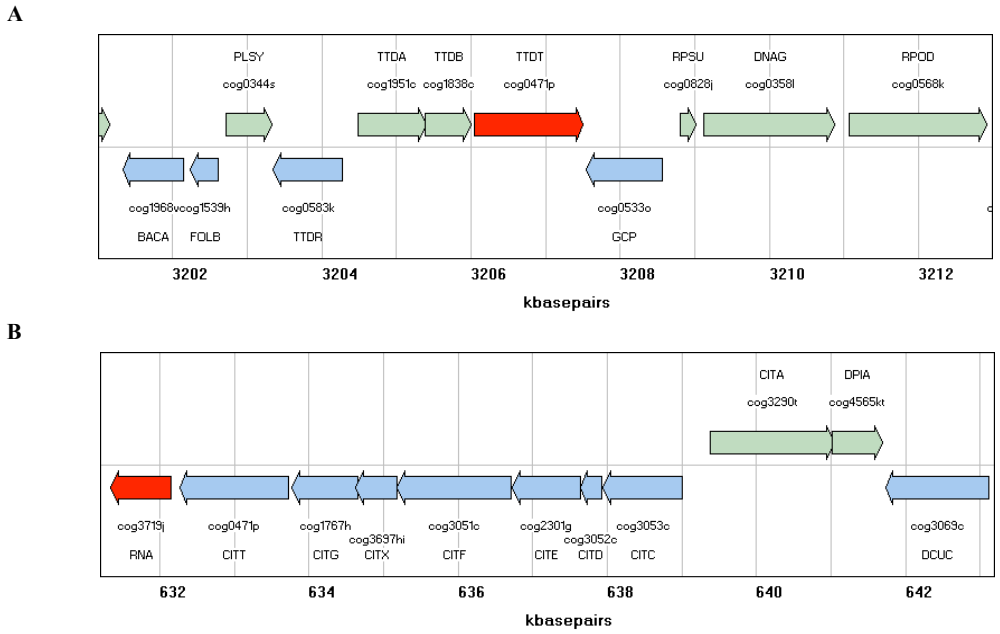


Fig. 9. Organization of the (A) L-tartrate and (B) citrate fermentation genes in *Escherichia coli*. See text for further details.

The gene encoding the regulatory protein ttdR is located immediately upstream and transcribed divergently. In case of the citrate pathway, the transporter gene *citT* clusters with the structural genes *citD*, *citE* and *citF* encoding the citrate lyase complex and the accessory genes *citC*, *citG* and *citX* that are involved in the maturation of the complex (Fig. 9B). A two component signal transduction system is encoded downstream of the operon in opposite direction and is most likely involved in the detection of citrate in the medium. The CitT transporter was cloned and functionally characterized as a citrate:succinate exchanger [36].

Involvement of TtdT in the metabolism of L-tartrate was first demonstrated by inactivation of the *ttdT* gene in a strain that was deficient in all known other C_4-dicarboxylate transporters [14]. The mutant strain was unable to ferment L-tartrate. The strain containing the active *ttdT* gene catalyzed uptake of L-tartrate, succinate and L-tartrate:succinate exchange. The present study confirms the catalytic activities of the TtdT transporter. Expression of the *ttdT* gene in *E. coli* SF100 resulted in cells capable of taking up tartrate and exchanging tartrate and succinate. Uptake of succinate could not be demonstrated in the genetic background of the SF100 strain. It follows that both CitT and TtdT catalyze a precursor:product type of exchange in which the uptake of the substrate, i.e. L-tartrate or citrate, is coupled to the excretion of the end product of the pathway (e.g. succinate).

4.2. Membrane topology of TtdT

The tartrate:succinate exchanger TtdT and the citrate:succinate exchanger CitT are homologous proteins and members of the DASS family in the Transporter Classification (TC) system. The DASS family is found in structural class ST[3] of the MemGen classification that also contains the well-studied 2HCT and ESS transporter families. Transporters in families in the same structural class of the MemGen system are believed to share the same global fold. The ST[3] fold features two homologous domains of 5 TMSs each (5+5 topology), oppositely oriented in the membrane (inverted topology), and a reentrant loop in between the 4^{th} and 5^{th} segment of each domain. The model of the 'core' structure is based on extensive studies of transporters in the 2HCT and ESS families [61,62]. Family averaged hydropathy profile alignment of the DASS and 2HCT families revealed the presence of an additional N-terminal TMS analogous to the one found in CitS of *K. pneumonia*, a member of the 2HCT family [61]. The MemGen membrane topology model for the TtdT and CitT transporters differs from models predicted by the secondary structure predictor TMHMM [70] that predicts 12 and 13 TMSs for TtdT and CitT, respectively. For both proteins, 7 TMSs are predicted in the N-terminal half with the N terminus in the cytoplasm and 5 or 6 TMSs in the C-terminal half. It was noted before that the presence of reentrant loop structures in membrane proteins is a source of errors in TMHMM models as these are not a 'state' in the Hidden Markov model and often erroneously are being mistaken for trans membrane segments. However, also a more recent version of a Hidden Markov predictor, termed Octopus [71], that does include reentrant loops in the model, does not recognize such features in the TtdT and CitT sequences. Octopus predicts the presence of 15 TMSs in both proteins. The MemGen and TMHMM models are significantly different and simple experimentation should be able to discriminate between them. Most prominent differences are the cellular disposition of the central loop, 'in' in the MemGen model and 'out' in the TMHMM model and the assignment of the hydrophobic region between positions 570 and 600 in the multiple sequence alignment (Fig. 3) that represents a reentrant loop in the MemGen model and a transmembrane segment in the TMHMM model. The cellular disposition of a set of 7 sites covering the C-terminal half of the TtdT protein determined by the complementary PhoA and GFP reporter fusion technique, clearly was in agreement with the MemGen model; the central loop resides in the cytoplasm and the

hydrophobic segment between the S396 and P455 positions (Fig. 4) does not span the membrane but rather loops back to form a reentrant or pore loop structure.

The DASS family is one of a few families in class ST[3] of the MemGen classification that contains transporters of eukaryotic origin. Well-studied are the Na^+-dependent dicarboxylate transporters NaDC1 and NaDC3 found in the apical and basolateral membranes of cells of renal proximal tubule, respectively. Topology models of these transporters are based on hydropathy profile analyses and secondary structure prediction algorithms, and reveal 11 to 14 TMSs [35]. The models were never tested thoroughly even though a significant body of SCAM (substituted-cysteine accessibility method) data is available. The topology model in use for the best studied transporter in the family, NaDC1, shows 11 TMSs with the N- and C-termini intra en extracellular, respectively. A number of residues, located in extracellular loop 5 and TMS 10 were sensitive to MTSET, a membrane impermeant thiol reagent, depending on the conformational state of the transporter (substrate protection) [72,73]. Sequence alignment revealed that the extracellular loop 5 and TMS 10 of the structural model of NaDC-1 corresponds to the pore loop region in the C-terminal domain of the MemGen model for the DASS transporters. An alternative model based on studies of human NaDC3 contains a reentrant loop between TMSs VIII and IX [74].

4.3. Accessibility of cysteine residues

The labeling studies with the thiol reagents FM and NEM using wild type and cysteine mutants of TtdT were inconclusive. Treatment of cells containing the cysteine mutants SCC, CSC and CCS with FM under conditions where the reagent reacts only with periplasmically exposed thiols, showed that Cys283 in the central loop was the only site that reacted with FM in the wild type protein, a result at variance with the cytoplasmic location of the central loop in the MemGen structural model. In fact, the result would support the TMHMM model (see above). However, the discrepancy with the MemGen model may be easily explained by the increased permeability of the membrane for FM caused by the expression of the TtdT protein. FM is able to react with cytoplasmic thiols when used at high concentrations [45]. Labeling by FM as an indicator of externally exposed thiols uses a concentration that was low enough not to show any labeling of

cytoplasmic thiols of the citrate transporter CitS of *Klebsiella pneumoniae* heterologously expressed in *E. coli* cells [45]. Under identical labeling conditions, the intensity of the FM fluorescent signal over the TtdT protein level clearly increased with increased induction strength (Fig. 6). The latter correlated with decreased tartrate and proline uptake, which is likely to be due to increased membrane leakiness. At $2x10^{-4}$ % of arabinose, the fluorescence over protein ratio is at most equal (Fig. 6,8) indicating that only a small fraction of the protein was labeled. Most likely, this is caused by a small amount of FM that leaks into the cell and reacts with Cys283 in the cytoplasmic loop of TtdT.

Treatment of wild type TtdT with NEM reduced the activity down to 40 %. Cys202 and Cys424 are located in the reentrant loops in the N and C terminal domains in the MemGen structural model of TtdT, respectively. Endogenous and engineered cysteine residues in reentrant loops of transporters from the 2HCT and ESS families in structural class ST[3] have been shown before to be the target site in similar inactivation studies. Alkylation of Cys398 or Cys414, both located in the reentrant loop in the C domain of the Na^+-citrate transporter CitS in the 2HCT family reduced the activity down to 15-20% [75]. Similarly, cysteine scanning experiments of the reentrant loops in the N and C domains of both CitS and the Na^+-glutamate transporter GltS revealed several sites that were highly sensitive to NEM treatment [62]. Unfortunately, the inactivation of TtdT could not be assigned to a single cysteine residue since replacing them one by one in the SCC, CSC and CCs mutants resulted in similar inactivation levels. One explanation might be that both cysteine residues in the two reentrant loops are labeled, but that only labeling one of the two is sufficient for the observed inactivation.

References

1. J.S. Lolkema & D.J. Slotboom, Sequence and hydropathy profile analysis of two classes of secondary transporters, *Mol. Mem. Biol.* **22(3)** (2005) 177-189.
2. G. Unden & J. Bongaerts, Alternative Respiratory pathways of Escherichia coli: energetic and transcriptional regulation in response to electron acceptors, *Biochim. Biophys. Acta* **1320** (1997) 217-234. Review
3. W.W. Kay & H.L. Kornberg, The uptake of C_4-dicarboxylic acids by *Escherichia coli*, *Eur. J. Biochem*. **18** (1971) 274–281.
4. I.G Janausch, E. Zientz, Q.H. Tran, A. Kröger, G. Unden, C_4-dicarboxylate carriers and sensors in bacteria, *Biochim. Biophys. Acta – Bioenergetics* **1553(1-2)** (2002) 39-56. Review.
5. S.J. Davies, P. Golby, D. Omrani, S. A. Broad, V. L. Harrington, J. R. Guest, D. J. Kelly, S. C. Andrews, Inactivation and regulation of the aerobic C_4-dicarboxylate transport (*dctA*) gene of *Escherichia coli*, *J. Bacteriol*. **181** (1999) 5624-5635.
6. K. Asai, S.H. Baik, Y. Kasahara, S. Moriya, N. Ogasawara, Regulation of the transport system for C_4-dicarboxylic acids in *Bacillus subtilis*, *Microbiology* **146(Pt. 2)** (2000) 263-271.
7. A. Kröger, S. Biel, J. Simon, R. Gross, G. Unden, C.R.D. Lancaster, Fumarate respiration of *Wolinella succinogenes*: enzymology, energetics, and coupling mechanism, *Biochim. Biophys. Acta* **1553** (2002) 23-38.
8. C.R.D. Lancaster & J. Simon, Succinate:quinone oxidoreductases from ε-proteobacteria, *Biochim. Biophys. Acta* **1553** (2002) 84–101.
9. C.J. Falzone, W.E. Karsten, J.D. Conley, R.E. Viola, l-Aspartase from *Escherichia coli*: substrate specificity and role of divalent metal ions, *Biochemistry* **27** (1988) 9089-9093.
10. W.E. Karsten, R.B. Gates, R.E. Viola, Kinetic studies of l-aspartase from *Escherichia coli*: substrate activation, *Biochemistry* **25** (1986) 1299-1303.
11. J.S. Miles & J.R. Guest, Complete nucleotide sequence of the fumarase gene *fumA* of *Escherichia coli*, *Nucleic Acids Res*. **12** (1984) 3631-3642.
12. R.H. Vaughn, G.L. Marsch, T.C. Stadtmann, B.C. Cantino, Decomposition of tartrates by the coliform bacteria*, J. Bacteriol*. **52** (1946) 311–325.

13. S.K. Reaney, C. Begg, S. J. Bungard, J. R. Guest, Identification of the L-tartratase genes (*ttdA* and *ttdB*) of *Escherichia coli* and evolutionary relationship with the class I fumarase genes, *J. Gen. Microbiol*. **139** (1993) 1523-1530.
14. O.B. Kim, G. Unden, The L-Tartrate/Succinate antiporter TtdT (YgjE) of L-Tartrate fermentation in *Escherichia coli*, *J. Bacteriol*. **189(5)** (2007) 1597-1603.
15. K. Fukui, C. Koseki, Y. Yamamoto, J. Nakamura, A. Sasahara, R. Yuji, K. Hashiguchi, Y. Usuda, K. Matsui, H. Kojima, K. Abe, Identification of *succinate exporter* in Corynebacterium glutamicum and its physiological roles under anaerobic conditions, *J. Biotechnol*. **154(1)** (2011) 25-34. / M. Inui, S. Murakami, S. Okino, H. Kawaguchi, A.A. Vertès, H. Yukawa, Metabolic analysis of *Corynebacterium glutamicum* during lactate and succinate productions under oxygen deprivation conditions, *J. Mol. Microbiol. and Biotechnol.* **7(4)** (2004) 182-196. / S. Okino, M. Inui & H. Yukawa, Production of organic acids by *Corynebacterium glutamicum* under oxygen deprivation, *App. Microbiol. and Biotechnol.* **68(4)** (2005) 475-480.
16. A. Böck & G. Sawers, Fermentation. In Escherichia coli and Salmonella: Molecular and Cellular Biology, 2nd Edition, (Eds. F. C. Neidhardt et al), *ASM Press* **18** (1996) 262-282.
17. M.E. Spencer & J.R. Guest, Regulation of citric acid cycle genes in facultative bacteria, *Microbiol. Sci.* **4** (1987) 164–167.
18. H.Q. Tran, J. Bongaerts, D. Vlad, G. Unden, Requirement for the proton-pumping NADH I dehydrogenase of *Escherichia coli* in NADH-fumarate respiration and bioenergetic implications, *Eur. J. Biochem*. **244** (1997) 155–160.
19. S. Six, S.C. Andrews, G. Unden, J.R. Guest, *Escherichia coli* possesses two homologous anaerobic C_4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct), *J. Bacteriol*. **176** (1994) 6470-6478.
20. P. Engel, R. Kramer, G. Unden, Anaerobic fumarate transport in *Escherichia coli* by an fnr-dependent dicarboxylate uptake system which is different from the aerobic dicarboxylate uptake system, *J. Bacteriol*. **174** (17) (1992) 5533–5539.
21. P. Engel, R. Kramer, G. Unden, Transport of C_4-dicarboxylates by anaerobically grown Escherichia coli. Energetics and mechanism of exchange, uptake and efflux, *Eur. J. Biochem.* **222(2)** (1994) 605–14.

22. E. Zientz, S. Six, G. Unden, Identification of a third secondary carrier (DcuC) for anaerobic C_4-dicarboxylate transport in *Escherichia coli*: roles of the three Dcu carriers in uptake and exchange, *J. Bacteriol*. **176** (1996) 7241-7247.
23. E. Zientz, J. Bongaerts, G. Unden, Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR*) two-component regulatory system, *J. Bacteriol.* **180** (1998) 5421-5425.
24. E. Zientz, I. G. Janausch, S. Six, G. Unden, Functioning of DcuC as the C_4-dicarboxylate carrier during glucose fermentation by *Escherichia coli*, *J. Bacteriol*. **181** (1999) 3716-3720.
25. P. Golby, D. J. Kelly, J. R. Guest, S. C. Andrews, Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C_4-dicarboxylate transporters in *Escherichia coli*, *J. Bacteriol*. **180** (1998) 6586-6596.
26. P. Golby, S. Davies, D. J. Kelly, J. R. Guest, S. C. Andrews, Identification and characterization of a two-component sensor-kinase and response regulator system (DcuS-DcuR) controlling gene expression in response to C_4-dicarboxylates in *Escherichia coli*, *J. Bacteriol*. **181** (1999) 1238-1248.
27. P. Dimroth & B. Schink, Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria, *Arch. Microbiol*. **170** (1998) 67-77.
28. B. Poolman, D. Molenaar, E.J. Smid, T. Ubbink, T. Abee, P.P. Renault, W.N. Konings, Malolactic Fermentation: Electrogenic Malate Uptake and Malate/Lactate Antiport Generate Metabolic Energy, *J. Bacteriol*. **173** (1991) 6030-6037.
29. M. Bott, Anaerobic citrate metabolism and its regulation in enterobacteria, *Arch. Microbiol.* **167** (1997) 78–88, minireview.
30. M. Lütgens & G. Gottschalk, Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate, *J. Gen. Microbiol*. **119** (1980) 63–70.
31. S. Kawai, H. Suzuki, K. Yamamoto, M. Inui, H. Yukawa, H. Kumagai, Purification and characterization of a malic enzyme from the ruminal bacterium *Streptococcus bovis* ATCC 15352 and cloning and sequencing of its gene, *Appl. Environ. Microbiol*. **62** (1996) 2692-2700.

32. S. Kawai, H. Suzuki, K. Yamamoto, H. Kumagai, Characterization of the l-malate permease gene (maeP) of *Streptococcus bovis* ATCC 15352, *J. Bacteriol*. **179** (1997) 4056-4060.
33. A.M. Pajor, Molecular properties of the SLC13 family of dicarboxylate and sulfate transporters, *Pflugers Arch.* **451(5)** (2006) 597-605.
34. K.M. Pos, P. Dimroth, M. Bott, The *Escherichia coli* Citrate Carrier CitT: a Member of a Novel Eubacterial Transporter Family Related to the 2-Oxoglutarate/Malate Translocator from Spinach Chloroplast, *J. Bacteriol*. **180(16)** (1998) 4160-4165.
35. M.H. Saier Jr, A functional-phylogenetic classification system for transmembrane solute transporters, *Microbiol. Mol. Biol. Rev*. **64** (2000) 354–411.
36. A. Weber, E. Menzlaff, B. Arbinger, M. Gutensohn, C. Eckerskorn, U-I. Flügge, The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells, *Biochemistry* **34** (1995) 2621–2627.
37. P. Golby, D.J. Kelly, J.R. Guest, S.C. Andrews, Topological analysis of DcuA, an anaerobic C4-dicarboxylate transporter of *Escherichia coli*, *J. Bacteriol*. **180(18)** (1998) 4821-4827.
38. J. Bauer, M.J. Fritsch, T. Palmer, G. Unden, Topology and Accessibility of the Transmembrane Helices and the Sensory Site in the Bifunctional Transporter DcuB of *Escherichia coli*, *Biochemistry* **50** (2011) 5925-5938.
39. C. Mulligan, M. Fischer, G.H. Thomas, Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archea, FEMS Microbiol. Rev. 35 (2010) 68-86.
40. R. ter Horst & J. S. Lolkema, Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity, *Biochim. Biophys. Acta – Biomembranes* **1818(1)** (2011) 72-81.
41. J.G. Shaw, M.J. Hamblin, D.J. Kelly, Purification, characterization and nucleotide sequence of the periplasmic C_4-dicarboxylate-binding protein (DctP) from *Rhodobacter capsulatus*, *Mol. Microbiol*. **5** (1991) 3055–3062.
42. J.A. Forward, M.C. Behrendt, N.R. Wyborn, R. Cross, D.J. Kelly, TRAP transporters: A new family of periplasmic solute transport systems encoded by the *dctPQM* genes of

Rhodobacter capsulatus and by homologs in diverse Gram-negative bacteria, *J. Bacteriol*. **179** (1997) 5482–5493.

43. R. ter Horst, J.S. Lolkema, Rapid screening of membrane topology of secondary transport proteins, *Biochim. Biophys. Acta* **1798(3)** (2010) 672-80.
44. R.D. Barabote , D.G. Tamang , S.N. Abeywardena , N.S. Fallah , J.Y. Fu , J.K. Lio , P. Mirhosseini , R. Pezeshk , S. Podell , M.L. Salampessy , M.D. Thever , M.H. Saier Jr., Extra domains in secondary transport carriers and channel proteins, *Biochim. Biophys. Acta*. **1758(10)** (2006) 1557-79.
45. T. Fujiki, K. Nanatani, K. Nishitani, K. Yagi, F. Ohnishi, H. Yoneyama, T. Uchida, T. Nakajima, K. Abe, Membrane Topology of Aspartate: Alanine Antiporter AspT from *Comamonas testosterone*, *J. Biochem.* **141** (2007) 85–91.
46. K. Nanatani, T. Fujiki, K. Kanou, M. Takeda-Shitaka, H. Umeyama, L. Ye, X. Wang, T. Nakajima, T. Uchida, P. C. Maloney, K. Abe, Topology of AspT, the Aspartate:Alanine Antiporter of *Tetragenococcus halophilus*, Determined by Site-Directed Fluorescence Labeling, *J. Bacteriol*. **189** (2007) 7089-7097.
47. K. Nanatani, P.C. Maloney, K. Abe, Structural and Functional Importance of Transmembrane Domain 3 (TM3) in the Aspartate:Alanine Antiporter AspT: Topology and Function of the Residues of TM3 and Oligomerization of AspT, *J. Bacteriol*. **191(7)** (2009) 2122-2132.
48. P. Dimroth, A. Thomer, Citrate transport in *Klebsiella pneumoniae*, *Biol. Chem. Hoppe Seyler* **367(2)** (1986) 813-823.
49. M.E. van der Rest, R.M. Siewe, T. Abee, E. Schwarz, D. Oesterhelt, W.N. Konings, Nucleotide sequence and functional properties of a sodium-dependent citrate transport system from *Klebsiella pneumoniae*, *J. Biol. Chem*. **267** (1992) 8971-8976.
50. Y. Wei, A.A. Guffanti, M. Ito, T.A. Krulwich, *Bacillus subtilis* YqkI is a novel malic/Na^+-lactate antiporter that enhances growth on malate at low protonmotive force, *J. Biol Chem.* **275(39)** (2000) 30287-30292.
51. B.P. Krom, R. Aardema, J.S. Lolkema, *Bacillus subtilis* YxkJ is a secondary transporter of the 2-hydroxycarboxylate transporter family that transports L-malate and citrate, *J. Bacteriol*. **183(20)** (2001) 5862-5869.

52. C.N. Kästner, K. Schneider, P. Dimroth, K. M. Pos, Characterization of the citrate/acetate antiporter CitW of *Klebsiella pneumoniae*, *Arch. Microbiol*. **177** (2000) 500-506.
53. C. Marty-Teysset, J.S. Lolkema, P. Schmitt, C. Divies, W.N. Konings, Membrane potential-generating transport of citrate and malate catalyzed by CitP of *Leuconostoc mesenteroides*, *J. Biol. Chem*. **270** (1995) 25370-25376.
54. D.J. Slotboom, J.S. Lolkema, W.N. Konings, Membrane topology of the C-terminal half of the neuronal, glial, and bacterial glutamate transporter family, *J. Biol. Chem*. **271** (1996) 31317-31321.
55. D.J. Slotboom, W.N. Konings, J.S. Lolkema, Structural features of the glutamate transporter family, Microbiol, *Mol. Biol. Rev.* **63** (1999) 293-307.
56. D.J. Slotboom, W.N. Konings, J.S. Lolkema, Cysteine-scanning mutagenesis reveals a highly amphipathic, pore-lining membrane-spanning helix in the glutamate transporter GltT, *J. Biol. Chem*. **276(14)** (2001) 10775–10781.
57. D. Yernool, O. Boudker, Y. Jin, E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, *Nature* **431** (2004) 811-818.
58. J.S. Lolkema, D.J. Slotboom, Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis, *J. Mol. Biol*. **327** (2003) 901-909.
59. I. Sobczak, J.S. Lolkema, The 2-Hydroxycarboxylate Transporter Family: Physiology, Structure, and Mechanism, *Microbiol. Mol. Biol. Rev*. **69** (2005) 665–695.
60. A.J. Dobrowolski, I. Sobczak, J.S. Lolkema, Experimental Validation of Membrane Topology Prediction by Hydropathy Profile Alignment: Membrane Topology of the Na^+-Glutamate Transporter of *Escherichia coli*, *Biochemistry* **46** (2007) 2326-2332.
61. F. Baneyx, G. Georgiou, In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT, *J. Bacteriol*. **172** (1990) 491-4.
62. L.M. Guzman, D. Belin, M.J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter, *J. Bacteriol*. 177 (1995) 4121-4130.
63. E.R. Geertsma, B. Poolman, High-throughput cloning and expression in recalcitrant bacteria, *Nature Methods* **4** (2007) 705-707.

64. C. Manoil, Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions, *Methods in cell biology* **34** (1991) 61-75.
65. D. Drew, D. Sjöstrand, J. Nilsson, T. Urbig, C.N. Chin, J.W. de Gier, G. von Heijne, Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis, *Proc. Natl. Acad. Sci. USA* **99** (2002) 2690–2695.
66. J.S. Lolkema, D.J. Slotboom, Estimation of structural similarity of membrane proteins by hydropathy profile alignment, Mol. Membr. Biol. 15 (1998) 33-42.
67. J.S. Lolkema, D.J. Slotboom, Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins, *FEMS Microbiol. Rev*. **22** (1998) 305-322.
68. A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol*. **305** (2001) 567-580.
69. H. Viklund, A. Elofsson, Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar, *Bioinformatics* **24** (2008) 1662-1668.
70. A.M. Pajor, K.M. Randolph, Conformationally Sensitive Residues in Extracellular Loop 5 of the Na^+/Dicarboxylate Co-transporter, *J. Biol. Chem*. **280(19)** (2005) 18728–18735.
71. A.M. Pajor, Conformationally Sensitive Residues in Transmembrane Domain 9 of the Na^+/dicarboxylate Co-transporter, *J. Biol. Chem.* **276** (2001) 29961-29968.
72. X. Bai, X. Chen, A. Sun, Z. Feng, K. Hou, B. Fu, Membrane topology structure of human high-affinity, sodium-dependent dicarboxylate transporter, *FASEB Journal* **21(10)** (2007) 2409-2417.
73. I. Sobczak, J.S. Lolkema, Alternating Access and a Pore-Loop Structure in the Na+-Citrate Transporters CitS of *Klebsiella pneumoniae*, *J. Biol. Chem*. **279(30)** (2004) 31113-31120.

Chapter 5

Summary and Conclusions

All living cells are surrounded by at least one biological membrane that forms a hydrophobic barrier between the interior of the cell and its environment. The uptake of nutrients, excretion of metabolic end products, removal of toxic compounds, maintaining osmolarity and internal pH require designated transport proteins that are embedded in the membrane. These integral membrane proteins are essential for selective solute transport and allow the cell to maintain homeostasis to ensure optimal conditions for metabolic processes. The importance of membrane proteins is further illustrated by the fact that about 20-30% of all genes in most genomes encode membrane proteins [1,2].

Chapter 2 shows how limited experimental data may be very useful to discriminate between membrane topology models of membrane proteins derived from different methods, such as TMHMM and MemGen. A membrane topology screening method is proposed by which the cellular disposition of three positions in a membrane protein are determined, the N- and the C-termini and a position in the middle of the protein. The method involves amplification of the encoding genes or gene fragments by PCR, rapid cloning in dedicated vectors by ligation independent cloning, and determination of the cellular disposition of the three sites using conventional techniques. The N-terminus was determined by labeling with a fluorescent probe, the central position and the C-terminus by the reporter fusion technique using alkaline phosphatase (PhoA) and green fluorescence protein (GFP) as reporters. The method was evaluated using 16 transporter proteins of known function from 4 different structural classes. For 13 proteins a complete set of 3 localizations was obtained. The experimental data was used to discriminate between membrane topology models predicted by TMHMM, a widely used predictor using the amino acid sequence as input and by MemGen, that uses hydropathy profile alignment and known 3D structures or existing models. It follows that in those cases where the models from the two methods were similar, the models were consistent with the experimental data. In those cases where the models differed, the MemGen model agreed with the experimental data. Three more recent predictors, MEMSAT3, OCTOPUS and TOPCONS showed a significantly higher consistency with the experimental data than observed with TMHMM.

The MemGen structural classification of membrane proteins groups families of proteins by hydropathy profile alignment. Class ST[3] of the MemGen classification contains 32 families of transporter proteins including the IT superfamily. In chapter 3, transporters from 19 different families in class ST[3] were evaluated by the TopScreen experimental topology screening method to verify the structural classification by MemGen. For nearly all transporters at least one of the predicted localizations is different in the models produced by MemGen and predictor TMHMM. Comparison to the experimental data showed that in all cases the prediction by MemGen was correct. It is concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other families in class ST[3]. The core structure of the model consists of two homologous domains, each containing 5 transmembrane segments, which have an opposite orientation in the membrane. A reentrant loop is present in between the 4^{th} and 5^{th} segment in each domain. Nearly all of the identified and experimentally confirmed structural variations involve additions of trans membrane segments at the boundaries of the core model, at the N- and C-termini or in between the two domains. Most remarkable is a domain swap in two subfamilies of the [st312]NHAC family that results in an inverted orientation of the proteins in the membrane.

The C_4-dicarboxylate transporters from the DASS, Dcu, DcuC and TRAP-T families, are found in the structural class ST[3] of the MemGen structural classification together with the ESS and 2HCT families. In chapter 4, the topological organization of the tartrate transporter TtdT from *Escherichia coli,* a member of the DASS family, was analyzed by in-frame translational fusions between a series of progressively truncated forms of the *ttdT* gene and downstream reporter genes (*gfp*, and *phoA*), and by cysteine accessibility studies. While the cysteine accessibility studies were inconclusive, the fusion data presented here showed that the core structure of the ST[3] model is adopted by TtdT. The L-tartrate:succinate exchange activity of the cloned TtdT transporter was verified, and the correlation between TtdT expression levels and transport activity was analyzed.

Summary for the unaquainted

All living cells are surrounded by at least one biological membrane that forms a barrier between the interior of the cell and its environment. The membrane that separates the cytoplasm (cell interior) from its environment consists of a bilayer of lipids that contain hydrophilic (water soluble) head groups and hydrophobic (lipid soluble) tails. The bilayer cannot be passed by most water-soluble compounds, and uptake of nutrients, excretion of metabolic end products, removal of toxic compounds, cell-to-cell communication, maintaining osmolarity and internal pH require designated transport proteins that are embedded in the membrane. A schematic representation of a cell is presented in Fig. 1.

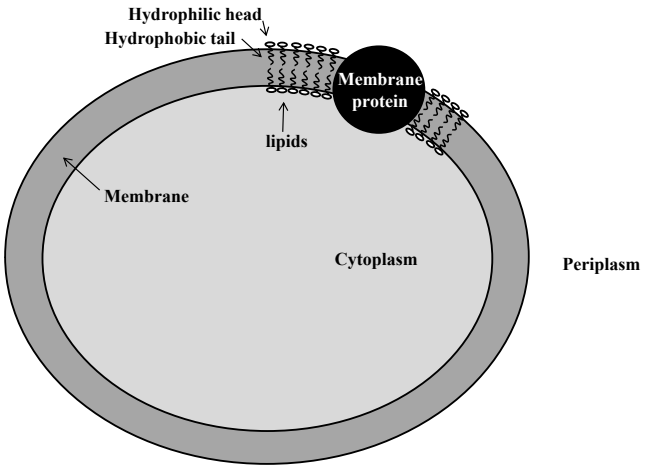


Figure 1. Schematic representation of a bacterial cell.

Based on certain characteristics (mode of transport, energy-coupling source), the Transport Classification (TC) system groups transport proteins in classes and subclasses. They are further subdivided in families and subfamilies, based on amino acid sequence identity and the compound(s) they transport. Members of a (sub)family are closely related, whereas proteins found in different families in a superfamily are more distantly related. Notwithstanding their great variety and number, α-helical membrane proteins share a common architecture: one or more bundles of membrane spanning hydrophobic regions connected by hydrophilic loops. The alternation of regions of high and low hydrophobicity give rise to a hydropathy profile characteristic to the membrane protein, and provides a fingerprint of the structure of the protein. In Fig. 2, a hydropathy profile of an α-helical membrane protein and a schematic representation of the corresponding membrane topology model is shown.

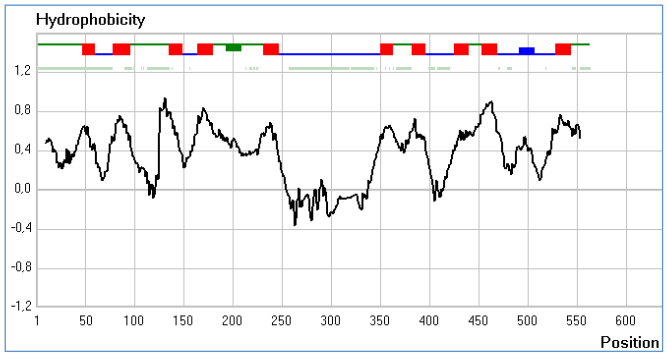


Fig. 2. Hydropathy profile and corresponding membrane topology model. Red squares, transmembrane segments (TMSs); green lines, external loops; blue lines, cytoplasmic loops; thickened loop regions (blue or green rectangles), reentrant loops.

In evolutionary terms, structure is in fact much better conserved than amino acid sequence. Consequently, classification of membrane proteins based on structure allows for the detection of more distant relationships than sequence alignment alone. In the MemGen classification system, similarities between (seemingly non-related) proteins or protein families are identified by comparison of hydropathy profiles. Each familie of membrane proteins is grouped in one of the structural classes, ST[1], ST[2], ST[3] and ST[4]. All members of a structural class share a global fold characteristic for that class. Although the MemGen classification system is not a membrane topology prediction method in itself, knowing the topology of one protein in a structural class infers knowing them all [33]. Structural classification by MemGen is supported by a number of high-resolution 3D crystal structures. At a lower level of resolution, structural similarities have been demonstrated for families in structural class ST[3], for which no resolution structures were available.

In Chapter 2, a membrane topology screening method (TopScreen) is proposed by which the cellular disposition of three positions in a membrane protein are determined: the head and tail (N- and C-termini) and a position in the middle of the protein. The locations of the C-termini of full-length proteins and of half-proteins were determined using reporter gene fusion techniques. Green fluorescent protein (GFP) and alkaline phosphatase (PhoA) were used as positive reporters of a cytoplasmic (inside) and periplasmic (outside) localization, respectively. GFP folds into an active conformation when localized in the cytoplasm, but not in the periplasm [28]. In contrast,

PhoA matures into the active state only when residing in the periplasm [29]. The third position, the N-terminus, was determined by the accessibility of an engineered cysteine residue using a membrane impermeable fluorescent reagent. The principles of the reporter and accessibility studies are represented in Fig. 3.

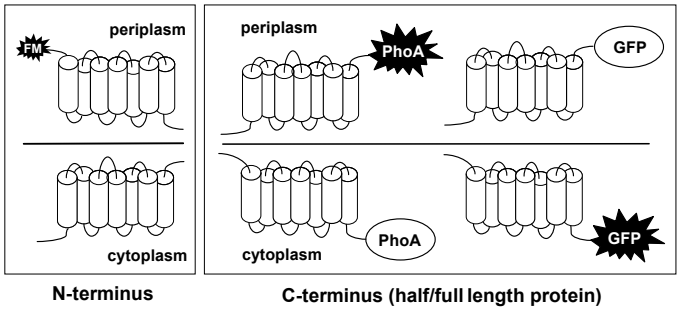


Figure 3. Schematic representation of the membrane topology screening. The N-terminal localization of the full-length protein is determined by the accessibility of the introduced cysteine residue at the external face of the membrane using fluorescent maleimide (FM), a membrane impermeable fluorescent sulfhydryl reagent. The C-terminal localizations of the full length and half proteins are reported by PhoA (periplasmic) and GFP (cytoplasmic) activity.

The method, TopScreen, was evaluated using 16 transporter proteins of known function from 4 different structural classes, and the experimental data was used to discriminate between membrane topology models from MemGen, the widely used topology predictor TMHMM and a number of more recently developed predictors.

In Chapter 3, transporter proteins from 19 different families in class ST[3] were evaluated by the TopScreen experimental topology screening method described in Chapter 2 to verify the structural classification by MemGen. As mentioned before, no resolution structures were available for families in structural class ST[3] of the MemGen classification. Class ST[3] contains 32 families of transporter proteins including the IT superfamily. It is concluded that the structural model available for transporters of the [st324]ESS and [st326]2HCT families is also valid for the other families in class ST[3]. A schematic representation of the core structure is shown in Fig. 4. The model consists of two homologous domains consisting of 5 transmembrane

segments each. The domains have opposite orientations in the membrane, and a reentrant loop is present in between the 4th and 5th segments in each domain.

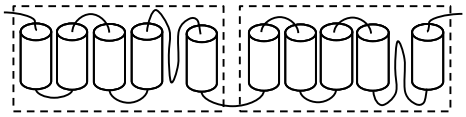


Figure 4. Membrane topology 'core' structure of transporters in structural class ST[3]. See text for explanation. Boxed parts represent domains. Each domain contains a reentrant loop between TMS 4 and 5.

Chapter 4 discusses the topological organization of the tartrate transporter TtdT from *Escherichia coli,* a member of the DASS family in structural class ST[3] of the MemGen classification, which was analyzed by in-frame translational fusions of progressively truncated forms of the *ttdT* gene and downstream reporter genes (*gfp* and *phoA*), and by cysteine accessibility studies. While the cysteine accessibility studies were inconclusive, the fusion data was consistent with the core structure of the ST[3] model. The L-tartrate:succinate exchange activity of the cloned TtdT transporter was verified, and the correlation between TtdT expression levels and transport activity was analyzed.

The objective of this research was to develop an experimental method that allows discrimination between topology models derived from different predictors. The method was subsequently used to experimentally evaluate the structural classification by MemGen. The TopScreen approach provided a large amount of data that was in full agreement with MemGen predictions, providing strong support for the typical fold of ST[3] transporter families. Discrepancies observed between TMHMM predictions and the TopScreen data most likely originates from the presence of reentrant loops. Based on the studies reported in this thesis, it was suggested that the 'core' structure identified in the [st324]GltS and [st326]2HCT families is also valid for the seemingly unrelated IT superfamily transporters and other families in class ST[3]. Nearly all of the identified and experimentally confirmed structural variations involve additions of transmembrane segments at the boundaries of the core model, the N- and C-termini or in between the two domains. Most remarkable is a domain swap in two subfamilies of the [st312]NHAC family that results in an inverted orientation of the proteins in the membrane.

Samenvatting

Alle levende cellen zijn omgeven door tenminste één biologisch membraan, welke een barriere vormt tussen het inwendige van de cel en haar omgeving. Het membraan dat het cytoplasma (cel interieur) scheidt van de omgeving, bestaat uit een dubbellaag van lipiden die hydrofiele (water oplosbare) hoofd groepen en hydrofobe (vet oplosbare) staarten bevatten. De dubbellaag is niet doorlaatbaar voor de meeste water oplosbare verbindingen. Een schematische weergave van een cel wordt gegeven in Fig. 1. De opname van voedingsstoffen, het uitscheiden van metabolische eindproducten, het verwijderen van giftige stoffen, cel-cel communicatie, en het handhaven van osmolariteit en interne pH zijn allen aangewezen op transporteiwitten die zijn ingebed in het membraan.

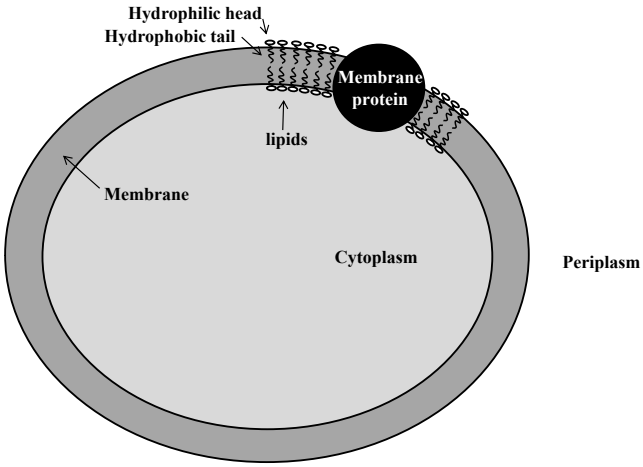


Figure 1. Schematische weergave van een bacteriële cel.

Het Transport classificatie (TC) systeem groepeert transporteiwitten in klassen en subklassen, op basis van bepaalde kenmerken (wijze van vervoer, energie-koppelingsbron). De klassen en subklassen worden verder onderverdeeld in families en subfamilies, gebaseerd op aminozuursequentie identiteit en substraat (de te transporteren verbinding). Transporters uit één (sub) familie zijn nauw verwant, terwijl de relatie tussen transporteiwitten uit verschillende families binnen een superfamilie soms niet eens detecteerbaar is. Ondanks hun grote aantal en verscheidenheid, hebben (α-helix) membraan eiwitten een gemeenschappelijk architectuur die bestaat uit één of meer bundels membraanspannende hydrofobe segmenten verbonden door hydrofiele lussen. De afwisselende gebieden van hoge en lage hydrofobiciteit leiden tot een hydropathie profiel karakteristiek voor het betreffende membraaneiwit. Het karakteristieke

hydropathie profiel is een soort vingerafdruk van de structuur van het eiwit. In Fig. 2 is een hydropathie profiel van een α-helix-membraaneiwit en een schematische weergave van het corresponderende membraan topologie model.

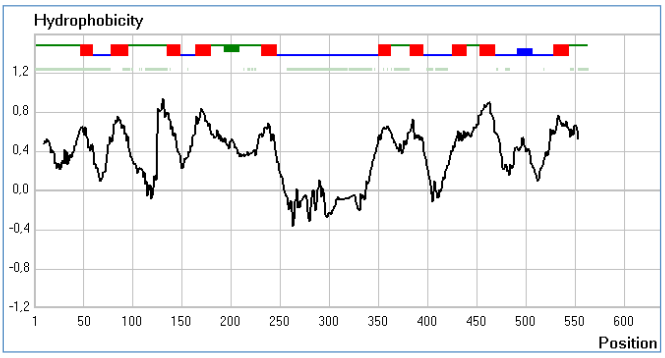
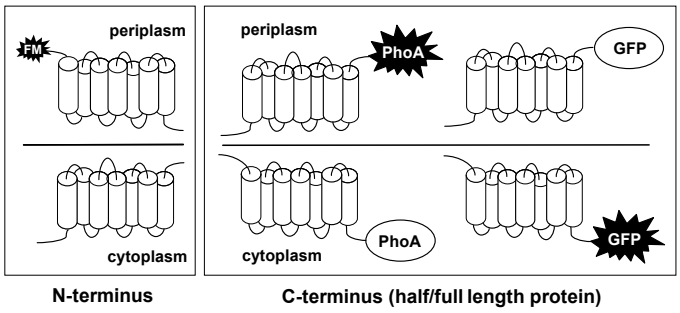


Fig. 2. Hydropathie profiel en corresponderend membraan topologie model. Rode vierkantjes, transmembraan segmenten (TMSen); groene lijnen, externe lussen; blauwe lijnen, cytoplasmatische lussen; verdikte gebieden in de lussen (blauwe of groene vierhoeken), pore loops.

In de evolutie, is structuur in feite veel beter bewaard dan aminozuur sequentie. Daardoor zal classificatie van membraaneiwitten op basis van de structuur in plaats van aminozuur sequentie leiden tot detectie van relaties op grotere afstand, dan in het geval van sequentie alleen. In de MemGen classificatie worden relaties tussen (schijnbaar niet-gerelateerde) transport eiwitten of families van transporteiwitten geïdentificeerd, door vergelijking van hydropathie profielen. Families van secundaire transporteiwitten zijn op deze wijze gegroepeerd in één van de structurele klassen, ST[1], ST[2], ST[3] en ST[4]. Alle leden van een structurele klasse hebben een globale structuur gemeen, die kenmerkend is voor de betreffende klasse. Dus, hoewel het MemGen classificatiesysteem in principe geen membraan topologie voorspellingsmethode is, kennis van de topologie van één eiwit in een structurele klasse, betekent kennis van de globale topologie van iedere transporter in die structurele klasse [33]. De MemGen classificatie wordt ondersteund door een aantal hoge-resolutie kristalstructuren. Op een lager niveau van resolutie, zijn structurele overeenkomsten aangetoond voor families in structurele klasse ST[3]. Tot op heden is er geen kristalstruktuur data beschikbaar van transporters in klasse ST[3].

In hoofdstuk 2 werd een membraan topologie screening methode (TopScreen) voorgesteld, waarin de locatie ten opzichte van het membraan van drie posities in het membraan eiwit wordt

bepaald: “kop” en “staart” (N- en C-termini) en een positie in het midden van het eiwit. De locaties van de uiteinden van volledige eiwitten en half-eiwitten werden bepaald met reporter gen-fusie technieken. Green fluorescent protein (GFP) en alkalische fosfatase (PhoA) werden gebruikt als positieve reporters voor, respectievelijk, cytoplasmatische (binnen) en periplasmatische (buiten) lokalisaties. GFP activeert wanneer de reporter zich bevind in het cytoplasma, maar niet in het periplasma [28]. Daarentegen wordt PhoA pas actief wanneer het zich bevindt in het periplasma [29]. Bij lokalisatie van derde positie, de N-terminus, werd er gekeken of een membraan-impermeabel fluorescerend reagens het N-terminale geïntroduceerde cysteine residu kon bereiken. De principes van de reporter en toegankelijkheid studies zijn weergegeven in Fig. 3.



Figuur 3. Schematische weergave van de membraan topologie screening. De N-terminale lokalisatie van het volledige membraaneiwit bepaald door de toegankelijkheid van het geïntroduceerde cysteïneresidu op het buitenvlak van het membraan met fluorescerent maleimide (FM), een membraan fluorescerende sulfhydryl reagens dat de membraan niet kan passeren. De C-terminale lokalisaties van de volledige en half-eiwitten bepaald door middel van reporteractiviteit van PhoA (periplasmatisch actief) en GFP (cytoplasmatische actief).

De methode TopScreen werd geëvalueerd met behulp van 16 transporteiwitten (met bekende functie) van 4 verschillende structurele klassen, en de experimentele gegevens werden gebruikt om te discrimineren tussen membraan topologie modellen van MemGen, van een veel gebruikte topologie predictor, TMHMM, en van een aantal recentelijk ontwikkelde topologie predictors.

In hoofdstuk 3, werden transporteiwitten van 19 verschillende ST[3] families beoordeeld met behulp van de experimentele topologie screening methode, TopScreen (beschreven in hoofdstuk 2), teneinde de MemGen structurele classificatie te verifiëren. Zoals eerder vermeld, zijn er geen

hoge resolutie structuren beschikbaar van ST[3] transporters van de MemGen classificatie. Klasse ST[3] bevat 32 families van transport eiwitten. Er werd geconcludeerd dat het structurele model beschikbaar voor transporters van de [st324]ESS en [st326]2HCT families ook geldt voor de andere families in de klasse ST[3]. Een schematische weergave van de kernstructuur is weergegeven in Fig. 4. Het model bestaat uit twee homologe domeinen, elk met 5 transmembraan segmenten. De domeinen hebben een tegengestelde oriëntatie in de membraan en in beide domeinen is er sprake van een pore loop, aanwezig tussen het 4e en 5e segment.

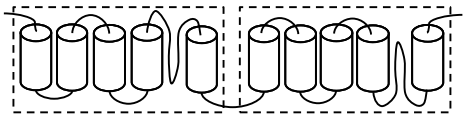


Figure 4. Membraan topologie kernstructur van ST[3] transporters. De met stippellijn omgeven gebieden zijn de domeinen. Elk domein bevat een pore loop tussen het 4e en 5e transmembraan segment.

Hoofdstuk 4 gaat in op de topologische organisatie van de tartraat transporter TtdT uit *Escherichia coli*, die behoort tot de DASS familie in structurele klasse ST[3] van de MemGen classificatie. De topologie van TtdT werd bestudeerd met behulp van reporter gen-fusie technieken, en cysteïne toegankelijkheid studies. Hoewel de cysteïne toegankelijkheid studies niet resulteerden in een overtuigend resultaat, waren de reporter gen-fusie resultaten in overeenstemming is met de kernstructuur van het ST[3] model. De L-tartraat: succinaat uitwisselingsactiviteit van de gekloonde TtdT transporter werd vasgesteld, en de correlatie tussen TtdT expressie en transport activiteit werd geanalyseerd.

Het doel van dit onderzoek was een experimentele methode te ontwikkelen die discriminatie mogelijk maakt tussen topologie-modellen van verschillende voorspellers. De methode werd vervolgens gebruikt om de structurele classificatie door MemGen te verifiëren. De TopScreen methode leverde een grote hoeveelheid gegevens die volledig in overeenstemming waren met MemGen “voorspellingen”, en de verschillen tussen de TMHMM voorspellingen en de TopScreen data zijn te verklaren met de aanwezigheid van de ‘pore loop’ regio’s. Op basis van de studies beschreven in dit proefschrift, werd gesuggereerd dat de kernstructuur van de [st324]GltS en [st326]2HCT families ook geld voor de schijnbaar ongerelateerde IT-superfamilie en andere

families in klasse ST[3]. Bijna alle geïdentificeerde en experimenteel bevestigde structuurverschillen omvatten toevoegingen van transmembraansegmenten rondom de domeinen in het model van de kern structuur. Het meest opmerkelijke was een ‘domain swap’, waargenomen in transporters van twee subfamilies van de [st312]NHAC familie. De ‘domain swap’ resulteerde in een omgekeerde oriëntatie van de eiwitten in het membraan.

Abbreviations

LIC	ligation independent cloning
GFP	green fluorescent protein
PhoA	alkaline phosphatase
FM	fluorescein-5-maleimide
NEM	*N*-ethylmaleimide
TMS	transmembrane segment
TA	tartric acid/tartrate
SA	succinic acid/succinate

Acknowledgements/Dankwoord

The completion of this thesis is probably the hardest thing I have ever done. The 4 or 5 years of research leading towards it, belong to the most exciting, challenging, stimulating, and fulfilling periods of my life. I can only hope that in the years to come I may find myself in likewise circumstances, to find the best and worst of who I am, so that life may never be dull.

Fortunately, I did not have to face my time as a Ph.D. researcher by myself. I have enjoyed the company of a great team of researchers from all over the world, broadening my field of interest not only in science but culturally as well. The magnificent variety in personalities gave an indispensable sparkle to my time in the Molmic group.

I would like to thank my promotor Arnold Driessen. Your continuous efforts in creating such a versatile and stimulating environment for Life Sciences have been crucial to many great accomplishments in scientific research. You have my gratitude for having been a part of your group.

Juke, I could not have wished for a better supervisor. The door to your office was always open. I have enormously appreciated your frequent visits in the laboratory, the scientific discussions, and your direct, no-nonsense attitude towards everything and everyone.

Adam, Tomek and Magda, these years would not have been the same without you. We have shared office, lab, summerschool, conference and party. Whenever I think of you guys, a big warm smile appears on my face.

Francesco, aka Dr. Bonny ‘Virtuoso’, despite the fact your time as a member of the Molmic group has ended a couple of years ago, we kept doing our lunch and coffee breaks until my very last days at the university. Our bond goes beyond my time as a Ph.D. researcher, and I hope it will last forever. You are a super gifted guitar player, and one of my best friends. May it be for life!

Alexej, you have a talent only few people posses. You perpetually radiate some kind of positive energy that made my occasional gloomy moods disappear as by magic. That, and of course the frequent barbeques you have organized, were a great addition to the enjoyment of life.

Bea and Manon, thank you so much for helping me out with the administrative work related to the completion of this thesis. Moreover, I have very much enjoyed your company. Molmic is lucky to have you!
Since I would like to stick to two pages of acknowledgments, I limit myself by thanking the rest of the Molmic group as a whole. Thank you all very much for a great learning and working environment, a very pleasant ambiance, your support, stimulating discussions and entertaining conversations. It has been a ball!

Martijn, Rogier en Wouter (niet noodzakelijk in deze volgorde ;), en vele anderen die ik gemakshalve niet bij naam noem, onze vriendschap is een onmisbare verrijking van mijn leven.
Pap, mam en Edwin, ik heb het jullie niet altijd even gemakkelijk gemaakt. Desondanks, heb ik altijd kunnen rekenen op jullie onvoorwaardelijke steun, vertrouwen en liefde. Al laat ik dat niet altijd blijken, ik ben jullie eeuwig dankbaar.
Stéphanie, my love, you have seen the best and worst of me, and stood by me at all times… You kept me going when I had (secretly) already given up. You have given me the greatest gift there is, our two sons, Bastien and Mattis. Whatever the future holds, as long as you, Bastien and Mattis are part of it, I have not much else to wish for. With all my heart, I will love you always.

List of Publications

R. ter Horst, J.S. Lolkema, Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity, *Biochim. Biophys. Acta.* **1818** (2011) 72-81.

R. ter Horst, J.S. Lolkema, Rapid screening of membrane topology of secondary transport proteins, *Biochim. Biophys. Acta.* **1798(3)** (2010) 672-80.

M.A. Kolbusz, **R. ter Horst**, D.J. Slotboom, J.S. Lolkema, Orientation of Small Multidrug Resistance Transporter Subunits in the Membrane: Correlation with the Positive-Inside Rule, *J. Mol. Biol.* **402(1)** (2010) 127-38.

E.P. Vos, **R. ter Horst**, B. Poolman, J. Broos, Domain complementation studies reveal residues critical for the activity of the mannitol permease from *Escherichia coli*, *Biochim Biophys Acta.* **1788(2)** (2009) 581-6.